DOCUMENT RESUME

ED 303 470 TM 012 213

AUTHOR	Dockrell, W. Bryan
TITLE	Achievement, Assessment and Reporting. Selected Essays. SCRE Publication No. 97.
INSTITUTION	Scottish Council for Research in Education.
REPORT NO	ISBN-0-947833-23-4
PUB DATE	88
NOTE	70p.
AVAILABLE FROM	The Scottish Council for Research, 15 St. John Street, Edinburgh EH8 8JR Scotland.
PUB TYPE	Collected Works - General (020) -- Reports - Research/Technical (143)
EDRS PRICE	MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	*Academic Achievement; Classroom Research; *Cross Cultural Studies; *Educational Assessment; Educational Policy; *Elementary Secondary Education; *Essays; *Foreign Countries; Graduation Requirements; Intelligence Tests; National Surveys; Personality Assessment; Policy Formation; Student Attitudes; Student Certification; Student Evaluation
IDENTIFIERS	Dockrell (W Bryan)

ABSTRACT

Five essays selected from 70 papers and books spanning 30 years in the publishing life of W. Bryan Dockrell are presented. The essays concentrate on three themes within the field of educational evaluation--attainment, assessment, and reporting. Perspectives and practices in different countries are provided. The essays include: (1) "The Contribution of National Surveys of Achievement to Policy Formation"; (2) "On Intelligence"; (3) "Assessment in the Classroom"; (4) "Reporting Assessments of Pupils' Attitudes and Personality"; and (5) "Certifying School Graduates." (TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED303470

ACHIEVEMENT ASSESSMENT AND REPORTING

Selected essays by

W. BRYAN DOCKRELL

DIRECTOR
THE SCOTTISH COUNCIL FOR
RESEARCH IN EDUCATION
1971-1986

TM 012 213

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

ROSEMARY WAKE

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Achievement, Assessment and Reporting

Selected Essays
by
W BRYAN DOCKRELL

Director
The Scottish Council for Research in Education
1971 – 1986

SCRE PUBLICATION 97

First published 1988

ISBN 0 947833 23 4

Copyright in this collection © 1987 The Scottish Council for Research in Education.
All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage or retrieval system, without permission in writing from the publisher.

Printed and bound in Great Britain for the Scottish Council for Research in Education by Russell Print, Forrest Street, Blantyre

CONTENTS

		page
Foreword		1
Introduction		3
1	The Contribution of National Surveys of Achievement to Policy Formation	7
2	On Intelligence	23
3	Assessment in the Classroom	31
4	Reporting Assessments of Pupils' Attitudes and Personality	41
5	Certifying School Graduates	49

Acknowledgments

'The contribution of national surveys of achievement to policy formation' first appeared in *Social Science Research and Public Policy-Making: a reappraisal* edited by D. B. P. Kallen, G. B. Kosse, H. C. Wagenaar, J. J. J. Kloprogge *and* M. Vorbeck, published by NFER-NELSON (1982). 'On intelligence' was first published in *On Intelligence contemporary theories and educational implications: a symposium, Toronto, 1969* edited by W. B. Dockrell, published by The Ontario Institute for Studies in Education (1970). 'Certifying school graduates' first appeared in *Evaluation Roles in Education* edited by Arieh Lewy *and* David Nevo, and published by Gordon and Breach in 1981. The Council gratefully acknowledges the permission given by these publishers to reproduce the papers in this volume.

Table 3.1 'Feedback from a diagnostic test on the environment' is reproduced from the table in *Diagnostic Assessment in Geography* by H. D. Black *and* W. B. Dockrell (1980), published by The Scottish Council for Research in Education. Table 3.3 'An extract from the test for use of the dative plural in German and the item rationale for the whole test' was first published in *Diagnostic Assessment in Secondary Schools* by H. D. Black *and* W. B. Dockrell (1980), published by The Scottish Council for Research in Education. Table 3.4 'Summary of pupil responses to a test for use of the dative plural in German' first appeared in *Diagnostic Assessment in Secondary Schools* by H. D. Black *and* W. B. Dockrell (1980), published by The Scottish Council for Research in Education. Table 5 'School leaving report - pupil profile assessment system', first appeared in *Pupils in Profile: making the most of teachers' knowledge of pupils*, The Scottish Council for Research in Education (author and publisher), 1977.

FOREWORD

Dr Bryan Dockrell

Dr W B Dockrell retired from the Directorship of SCRE on 1st August 1986. Professor John Nisbet, Chairman of the Council from 1975 to 1978, writes of his career:

Bryan Dockrell was one of the Godfrey Thomson's graduates in Edinburgh in the 1950s. After a spell in England, he moved to Alberta as Associate Professor of Educational Psychology. He took his doctorate in Chicago, and for five years before returning to Edinburgh he was Professor of Special Education in Ontario.

Scottish Council for Research in Education was entering a period of radical change when Dr Dockrell was appointed Director in 1971. Previously much of the Council's work had been done by committees giving their time to research on a voluntary spare-time basis. The new Director's task was to build a team of full-time professional researchers within a new structure of Scottish Education Department support and a programme of policy-oriented projects. By 1979-80 he had recruited (and gained funding for) a staff of no less than 47 researchers and support personnel.

Dr Dockrell's main contributions to research have been in the field of educational assessment: he was a member of the Dunning Committee and of the world-wide IEA project, and his international contacts brought many distinguished scholars to Scotland to give lecturers and seminars for SCRE. His work on pupil profiles is widely acknowledged, perhaps more outside Scotland than here at home. Within SCRE, one of his major achievements has been to introduce new styles of research which help to bridge the gap between researcher and practitioner.

Bryan Dockrell's 15 years at SCRE covered a period of unprecedented growth in educational research, and Scottish education is indebted to him for his important part in that development.

INTRODUCTION

It is no easy task to encapsulate more than 30 years of distinguished experience as a teacher and an educational researcher in a small collection of papers such as this. It is made more difficult when Bryan Dockrell's publications include more than 70 papers and books across a substantial portion of the educational spectrum. In choosing the papers for this collection I have, therefore, concentrated on three themes which are very much the concern of education today. These themes - attainment, assessment and reporting - are areas in which Bryan Dockrell's contribution is widely and internationally acclaimed.

Attainment

The first two papers focus on attainment, albeit from very different perspectives. The first, 'The Contribution of National Surveys of Achievement to Policy Formation' is important because it raises fundamental issues about the potential of surveys. It also provides an historical insight into the use of surveys of achievement by politicians and policy-makers. Drawing on the 1953 and 1963 *Scottish Scholastic Surveys*, the paper asks whether there was evidence that they did, in fact, influence policy-making at both national and classroom levels. Thus, for example, gains in achievement between the 1953 and 1963 tests were greater in areas where the local authority continued to make use of a battery of attainment tests on the completion of primary schooling; there was a positive correlation between high English scores and primary schools with libraries; pupils from smaller schools attained virtually the same standards as those in larger schools.

And the consequence for policy formulation? Attainment tests at the end of primary schooling have been largely abolished, and Dockrell could find no evidence that the findings had been used either to support the argument for library provision, nor in the still raging debate about school closure. The paper clearly recognises that the policy-makers may for other reasons have been right, but the disturbing argument is that the evidence from the surveys would appear to have been ignored rather than included in the justification of policy.

The discussion of impact on classroom policy is no less interesting. The question is: what should a teacher *do* with the information arising from such surveys?

> How would a particular teacher know whether the greater attention which, it is held nationally, should be paid to the layout of a short division sum, applies to his class? If he was already providing more attention than the average, should he provide not more but perhaps less? Is it likely that those already giving 'considerable attention to the layout of short division sums' would feel strengthened in their conviction, and provide even more? Would those not giving sufficient attention have overlooked this point in the recommendations of the report?

The curricular and pedagogic arguments on which the first paper is based clearly relate to discrete domains of subject-specific attainment. But this, of course, was not the psychometric tradition of large-scale testing, which was built on widely accepted assumptions about 'general attainment' or 'intelligence'. Today, almost all self-respecting educationists would admit to at least scepticism about the concept. But when Bryan Dockrell edited *On Intelligence*, the second essay in this collection, following a symposium in Toronto in 1969, the debate was of much greater immediacy. The paper is clearly different from the other papers in this volume because it itself formed the introduction to a series of papers. But as the contributors to the symposium included such interesting names as Cyril Burt, Arthur Jensen and Philip Vernon, no attempt has been made to alter this synthesis of the complex arguments.

Assessment

While the first two papers owe much to Dockrell's background as an educational psychologist, the next illustrates how he used this to advantage in dealing with contemporary problems of schools and classrooms.

'Assessment in the Classroom' has its roots in SCRE's work on profiling in the 1970s. Working with Patricia Broadfoot, and with the inspiration and support of both the Head Teachers' Association of Scotland and a senior member of HM Inspectorate, Dockrell produced *Pupils in Profile*, which most would agree to have been seminal in promoting thinking about profiling both within the UK and throughout the world. However, the original project left difficulties in two areas: how to deal with personality and attitudinal characteristics which are so difficult to assess, and what should be the nature and the function of subject-oriented assessment in the classroom.

Bryan Dockrell sat on what has come to be known as the 'Dunning Committee', which, between 1975 and 1977, considered what might be the most effective forms and functions of assessment and certification for 14 - 16 year-olds in Scottish schools. The Report was important not only because it established a strategy for certification which would meet the needs of *all* young people in their age range, but because it broadened the legitimate concern of assessment to include many more purposes than the summative. In 'Assessment in the Classroom', we have an account of how Dockrell came to interpret and to evaluate this notion of 'diagnostic assessment' which, although not new, was a novelty for many of the teachers who encountered it first through the work of SCRE.

Reporting

Assessment, perhaps, inevitably leads to reporting. Not least difficult is the question of what aspects of affective attainment might be reported to parents and, at a later stage, to employers. In 'Reporting Assessments of Pupils'

Attitudes and Personality' Bryan Dockrell draws on a series of studies carried out at SCRE to consider whether such characteristics should be reported. He then reports findings on the views of parents, teachers and young people, and concludes that while this is clearly an area of controversy, there can be no denying that, if these assessments are going to be made, they must be made well. Few would question this, but many would doubt whether its implications are widely understood.

The final outcome of the educational process in most societies is the leaving certificate, and in 'Certifying School Graduates' the final paper offers an interesting account of how this is dealt with in a number of countries. Despite its formal status as a simple record of achievement, the conferring of the school certificate has almost ritual status. Furthermore, it may be used (or demanded) as an indication of achievement for many years after its award. Dockrell explores the various conceptualisations of certification held by parents, employers and the young people themselves. He looks at the ways in which teachers use the system both as a stick and a carrot. He notes that the forms of certification and the grounds on which they are awarded vary substantially from one country to the next. Nevertheless, the award of some form of leaving certificate is a universal or near-universal phenomenon, and the practices of each country seem remarkably impervious to change.

Harry Black
The Scottish Council for Research in Education, August 1987

1

The Contribution of National Surveys of Achievement to Policy Formation

This paper was prepared for the International Workshop on Educational Research and Public Policy-Making held in May 1981 and organised by the Foundation for Educational Research in the Netherlands (SVO) under the auspices of the Secretary General of the Council of Europe.

The issue for this paper is how educational research can contribute to the formation of policy. There are those who argue that the primary contribution of research is to the shaping of the climate of opinion. Certainly much research makes its impact by its contribution to the consensus, to the general feeling that exists within the informed community. There are however, questions to be asked about this argument. The first question is whether research findings do in fact contribute to the climate. Are they simply used if they happen to fit the existing climate and ignored or forgotten if they do not? A second question is whether research which appears to make an impact does so, or whether it is the climate of opinion which determines the interpretation that the researchers place on their findings. Research findings do not exist in abstract. They are the constructs of researchers. The researcher sees through a filter, through a set of expectations. Research which is cited as having contributed to the climate of opinion consists often of conclusions in harmony with the existing presuppositions of the author and a climate of opinion which may not be general but is that of an intellectual élite.

This particular approach has become popular with the growing disenchantment that we see on both sides of the Atlantic with the contribution that research and evaluation can make to specific policy decisions. It may be a strategic withdrawal to previously prepared positions: a safe retreat for the academic. It is not however a satisfactory answer for the politician or administrator who is asked to provide millions of pounds or dollars or guilders. He may legitimately feel that he wants more for the public's money than that.

The contribution of this paper to the discussion is a presentation of a case study. The example of research selected is a national survey of achievement. There are several reasons for choosing this particular example as a case study. The first one is that so much time, effort and resources are being devoted internationally to surveys. National or regional surveys are being carried out or planned in many countries. The best known are the American National Assessment of Educational Progress (NAEP) studies and the state-wide assessment programmes in a number of American states. In Europe a number of such studies have been launched in England and others are being planned or discussed elsewhere. The last round of International Association for the

Evaluation of Educational Achievement (IEA) surveys involved 26 nations (Walker, 1976). The next round of surveys which is currently being planned will involve many more though the exact number is not yet certain. The survey therefore is an example of one kind of prominent research activity.

The second reason for choosing it was that the particular surveys discussed were examples of good educational research. That is, they were carefully planned and meticulously carried out. The third reason is that the studies did produce valid findings. Much research of all kinds including educational research leaves us very little wiser than we were before we began. That is not something to be surprised at, or something to be concerned about. It is to recognize the limitations of the human endeavour. In this particular case however, valid findings were produced.

Another reason for choosing these studies is that their findings had relevance at various levels: relevance to administrators, to teachers and to parents. A fifth reason for selecting these studies is that their findings are relevant to general issues and not simply to particular local and temporal questions. If it is argued that good research will inevitably contribute to general thinking then these studies can serve as an empirical test for that hypothesis. Finally, these surveys are important because they show the advantages and limitations of survey work, what can be learned from surveys and what cannot.

The Scottish system

The Scottish educational system is like the English system in some respects, in that it divides responsibility between the central agency, the Scottish Education Department (SED), and the local education authorities. The system is described in the booklet the *Educational System of Scotland* (SED, 1977, p21) which states that the education authorities 'are required to ensure that there is adequate and efficient provision of school education for children in their areas... They are responsible for the curriculum taught in their schools, head teachers normally exercising that responsibility on their behalf'.

Central Government, on the other hand, 'generally oversees the planning of school provision by education authorities and matters such as staffing, curricula, teaching methods, equipment, attendance and support of pupils... [it also] prescribes the requirements for entry to teacher training on the advice of the General Teaching Council for Scotland'.

There is therefore a division of responsibility between a central authority which has a supervisory and a guiding role and local authorities which are responsible for the context of the curriculum and the methods of teaching. There is no centrally prescribed curriculum, no list of approved textbooks. Such a system permits a great deal of diversity and indeed there is a substantial variability among Scottish primary schools. In these circumstances it is more difficult to monitor standards of achievement than in more centralized systems

where expectations are more precisely defined. Nevertheless it was believed that there was sufficient consensus for generally applicable tests to be devised and administered.

The 1953 Scottish Survey

There had been previous nationwide surveys involving the application of intelligence tests in 1932 and in 1947, and it was accepted that valuable information had been obtained about variations in intelligence. As the first report on the surveys, *The Scottish Scholastic Survey 1953* (SCRE, 1963, p17) states: 'It was thought that equally useful knowledge about the spread of the scholastic attainment of pupils could be found from the results of a similar national survey involving educational tests'. Among the useful knowledge that it was expected that the survey would gather was information about 'the amount of acceleration and retardation in the schools system (that is, grade skipping and grade repetition). The relative educational standards for urban and rural schools and of different sizes of schools and of schools organised on individual as compared with class methods'. If there were more specific objectives than these for the survey they are not stated in the report.

Tests of arithmetic (mechanical and reasoning) and English (usage and comprehension) were administered to 76,121 ten-year-olds (all those born between the 1st of July 1942 and the 30th June 1943) except those 'thought by their teachers to be unable to tackle with any hope of success tests primarily designed for normal pupils in their age group' (p21). The report explains why the particular age group was chosen, records the tests and reports in detail the sample.

It was not thought necessary to justify the content of the test except in the most formal sense. The curriculum as it existed was taken as a given. 'The tests set were restricted to what was assumed to be common to the schemes of work of all the areas' (p83). There was none of the careful sifting of aims and content that now takes place. Attention was however given to an issue that still is contentious. In order to overcome 'the fear that comparisons between the results of schools or of separate authority areas would be made as a result of the investigation, an assurance... [was] given that the survey results will be published in a form in which no such comparison would be possible' (p18).

The results were given in full and informative detail and a substantial number of conclusions were drawn. 'In the first place it has been shown that a scholastic survey on a national scale is possible' (p185). That in itself is important since nearly 30 years later there are still a number of countries where no national survey has been attempted and where there is considerable doubt about the feasibility of such surveys. 'The survey has also shown the difficulties of which the principal one is the diversity of work normally professed by an age group. At the ten-year-old level chosen for the survey this was particularly evident in

the subject of arithmetic where the complicated British tables of money, length and weight were introduced in different ways at different times in different areas' (p185).

The report goes on to say wisely that 'it will be folly to attempt to standardise curricula in this field until it has been shown that one method is superior to others' (p185). This quotation highlights two issues for those considering national surveys today. The first is the great variation in test score which reflects not long-term differences in level of attainment but short-term consequences of different teaching methods. The second is the danger of a backwash in the schools. If there are standard assessments which are administered nationally it may be assumed that these define a national curriculum. Even in a decentralised system schools will be under pressure to adopt this putative national curriculum. This is a fundamental issue which is discussed in more detail later.

A number of general conclusions were drawn. The usual sex differences were noted: 'boys and girls attained approximately the same standards in mechanical arithmetic while the boys were superior to the girls in arithmetical reasoning. In both tests of English the girls were superior to the boys...' (p151).

'The association between tests score and type of area (city, large town, small town, rural) was slight and for practical purposes the average performance by pupils in each of the four types of area was the same' (p155). A similar conclusion was drawn about differences in the ten regions: 'While there are variations in the attainment in the four tests the total attainment does not vary greatly from region to region' (p158). Nonetheless an inspection of the data indicates that scores from pupils in the Edinburgh and Dundee areas were generally high and the scores of those from the Glasgow area were generally low.

A careful analysis was made of the performance of left-handed children and the conclusion that was drawn was that the superiority of the right-handed group was probably a real one but was so slight as to be insignificant from the educational point of view.

The report turned next to the question of class size. It was careful to draw attention to the various factors that might be involved and rightly concluded that 'it will be apparent that there is no regularity about the results... It does not follow that size of class has no effect on attainment. The conclusion to be drawn is rather that it will be difficult to obtain definite conclusions on this topic with an experiment which is not specifically designed for the purpose' (p162).

On the impact of school size an issue which is still relevant in the United Kingdom and doubtless elsewhere the conclusion was 'the performance of the pupils in these schools (smaller) was on the whole as good as that of pupils in larger schools. In particular pupils from one teacher schools reached the same

standard as those attained by pupils in schools with more than six teachers' (p168).

The difficulties which the authors had pointed to in drawing conclusions about curriculum (referred to above) did not deter them.

The panel dealing with arithmetic arrived at very specific conclusions:

> Division by factors is undesirable in the primary schools...
> More attention should be paid to the lay-out of short division sums...
> There is need for standardising the notation used in recording the time of day by the clock. The panel recommends that for written expression it should be in the form of 8.50 a.m. Use of written working in arithmetic facilitates accuracy. Further use of working is helpful to a teacher in diagnosing a pupil's difficulties.
> A standard practice is required for recording remainders in division...
>
> The final point...question of the use of English. It was evident that the various aspects of teaching arithmetical problems required further consideration e.g. the need for accurate reading of the question and for noting units used (p186).

The panel dealing with the English tests arrived at equally definite conclusions:

> The tests in English usage demonstrated the need for persistent oral practice in accepted speech forms and a restrained use of pencil and paper exercises for occasional testing...
>
> Reading as a thought-getting process seemed insecure. It is possible that acquaintance with forms of verbal testing and the common use of reading textbooks with exercises make it all too easy to suppose that pupils working through a series of questions have understood what they are reading. The tests in this survey showed unmistakeably that many pupils dealing as well as they could with details have not first grasped the general meaning of what they had read (p187).

A survey which had begun with primarily structural objectives had been used to draw mainly curricular conclusions.

The 1963 Scottish Survey

The second survey was reported in *Rising Standards in Scottish Primary Schools* (SCRE, 1968). The objectives were apparently no more detailed than those of the earlier one. The report simply states that it was decided to conduct a second survey because 'it was hoped that besides indicating any changes in attainment that might have taken place in the ten intervening years a new survey might give some indication of the possible effects of new teaching methods' (p17).

The same tests were used as had been used on the previous occasion. Apparently the earlier tests were thought to be entirely satisfactory since they

were not revised at all. The test booklets were surplus stock from the 1953 survey. On this occasion however, a stratified random sample of 5,209 pupils was tested, not the whole age group.

The answer to the question which was the basis of the second survey had been given in the title of the report, *Rising Standards in Scottish Primary Schools*. 'Between the 1953 and 1963 surveys the changes in score in each of the four tests have been in an upward direction. The sizes of the gains are about one-third of the standard deviation of the distribution of scores or roughly the gains that will be made in six months by an average ten-year-old pupil' (p85). The study however, looks not only at general differences. The changes are related also to levels of ability, sex, types of area, region, sizes of schools, aspects of the tests and so on. 'The gains have been made by pupils at all levels of ability, by boys and girls to the same extent, in all regions of the country and in all sizes of schools... while performances on some items show greater improvement than on others, the gains have been spread over nearly all of the items of the test. They are attributable partly to greater speed in response and partly to greater accuracy when the responses have been made' (p85). The researchers dismiss test sophistication as a possible cause of these changes.

They then go on to look at specific instructional and administrative arrangements. The retention of attainment tests (previously used universally for selection for secondary education), the use of the Cuisenaire method, provision of libraries in schools, the effects of shortage of teachers, and left-handedness. Their conclusions on these issues vary. 'Areas still using attainment tests at the transfer stage show gains about twice as large as those in other areas' (p85). 'Little or no association was found between attainment in the arithmetic test and the use of Cuisenaire methods' (p85). 'Higher attainments in the English test go with a greater provision of school libraries' (p85). They go on to point out cautiously 'a cause and effect relationship cannot be assumed; both of these results could be due to a common third factor' (p85). One wonders equally whether a common factor could not have been responsible for the relationship between use of attainment tests and achievement. On the question of the shortage of teachers they conclude 'no association has been found between attainments and the shortage of teachers' (p85), but again caution rules the day and the report points out 'the sample data provide only scanty information on this point' (p85).

The committee were hesitant however, to make the same kinds of comments on teaching as had been made ten years earlier. For the most part they simply drew attention to the items where there had been changes and those where there had not. A few points however, were made. 'Computational errors still persist. Fractions are still being treated by some pupils by rote and long division is still insecure. The concept of zero as a place-holder is unfamiliar to many pupils' (p128).

In English some of the deficiencies noted ten years earlier were less conspicuous. 'Pupils were reading with more skill and becoming more independent in their thinking about what they read' (p128). They could not however, resist drawing special attention to a specific point. 'A disappointing feature for Scots was that the Scots poem showed the least gain of any section. Printed Scots is becoming completely unfamiliar to Scottish children' (p128). (It remains completely unintelligible to English adults.) In the detailed comments on the responses to the Scottish poem it was noted that Scots words were becoming even less familiar than they had been ten years earlier. 'Kye might as well have been a foreign word. The popular error was key followed by sky. Other suggestions (most of them not unreasonable) were pigs, horses, sheep, crops, corn, wheat, hay, children, keys' (p100). The authors go on to comment perhaps despairingly that 'All right" may have been an interpretation of "O'Kay' (p100). 'To the majority the Scots forms were not intelligible, and from the errors in other words not dialectical it was obvious that a large number did not begin to understand what the lines were about. Neither did they have the benefit of hearing them read or spoken. The 1953 comment is reiterated. "When one considers the extent to which Scots of some kind is spoken and understood one can only conclude that Scots in print is completely unfamiliar to three-quarters of the pupils in this age group...it would appear desirable to include some printed Scots among the reading material for Scottish children"' (p100).

The surveys were carefully designed, meticulously conducted, reported comprehensively and many conclusions relevant to policy were given. What impact did they have?

General issues

Before turning to that question which is the major one for this paper, it is worth noting some issues which were given less attention then than they would be given now.

In more recent studies, the National Assessment of Educational Progress in the United States for example, more thought has been given to the content of the tests than in Scotland or at least more of the effort put into deciding the content of the test has been recorded. Objectives are defined which have to be acceptable to the subjects' specialists, teachers and thoughtful adults. The items are chosen not to spread those taking the examination for selection purposes but rather are intended to indicate what proportion of the age group has mastered a particular aspect of the subject.

As with all criterion-referencing there is a problem of validity and in this case content validity is determined by a lengthy process of review involving the three groups of specialists referred to above. The National Assessment results indicate the proportion of the age group reaching the pre-defined criteria.

These results are reported in lengthy bulletins which are prepared by the NAEP which attempt to interpret the meaning of tests results and not simply to report them.

The scope of the Scottish tests was limited to what could be accomplished in two and a half hours. There was not matrix testing such as Carlson (1980) describes in California where the complete battery consists of 1,020 items. 'The long test battery means that it is possible to assess a much wider array of skills and concepts than would otherwise be possible' (p14). In the Scottish survey, as noted above, tests were restricted to what was assumed to be common and could not cover the many alternatives of content and method that can be covered in California.

Nor were there any attitudinal measures. The Scottish survey could only show what pupils in different sizes of schools achieved, not what attitudes were developed. Did school libraries result in more extensive reading and greater pleasure in reading as well as in higher attainment? The Scottish pioneers did not set out to gather such data.

Even more fundamental questions were left unanswered. The first task in a survey is to define the aspects of school work which are to be assessed. At the primary level should assessments be related to the traditional division of arithmetic and English as in the Scottish survey or should they be interdisciplinary and focused on the child's ability to solve problems drawing on all the experiences that the school provides? Do we want to know whether a pupil has acquired the basic skills allowing him to tackle particular problems or do we want to know whether he has also learned to apply the skills in a realistic situation?

There may be a sharp distinction between the words a child can decipher, those which he can interpret and those which he can use. In arithmetic there may be a gap between a pupil's ability to recite number facts and to use his understanding of those numerical relationships.

The Scottish pioneers did not ask as we would today, why do we teach children arithmetic or reading? What effects do we expect them to have on pupils when they have become adults? Are we moving to a society where the standards required of a minority will be far in excess of what we conventionally defined as literacy or numeracy, and relate to an ability to absorb complex ideas presented in a variety of media and to the ability to think mathematically about a range of problems? And, where the standards required of the majority will be limited to the ability to find Page 3 of the *Daily Mirror* and to calculate the stake money for a football pool entry?

Appropriate standards required in a future society have to be defined and this is not an issue that can be burked by taking refuge in the use of established tests. What applied to tests of English and arithmetic in the Scottish studies

applies to surveys of science and social studies. Do we merely wish children to be able to reproduce a series of facts, formulae and theories or to understand the scientific method? Is there a purpose to teaching about the Battle of Bannockburn, if so what is it and how can it be assessed?

Impact

When we turn to the question of impact we must first ask what we can expect such exercises to achieve.

As noted above the objectives specified for the Scottish Scholastic Surveys were very limited. What information was sought is defined but the use that could be made of it is not. A later and considerably more detailed statement is made in a leaflet, *Why, What and How*, produced by the English Assessment of Performance Unit (1977). The purpose of monitoring, it says, is to provide national information, not only to describe the current position but also to record changes as they occur. Further, such information would help to determine policy, including the making of decisions about the employment of resources. It would also help teachers in planning the balance of pupils' work in schools, without attempting at national level to define detailed syllabus content. Moreover, the outcomes of the tests were expected to make parents, employers and other concerned bodies better informed about the achievement of schools.

There are three sets of objectives: to provide information about matters of general policy, to provide information for teachers and to make parents, employers and others better informed. The Scottish Mental Survey provided information relevant to each of these objectives. I will look at the recommendations that refer to each of these issues in turn.

General policies

Much of the information at the national level was primarily of negative value. The differences among pupils in different types of areas (cities, large towns, small towns, other areas) could be dismissed. There was therefore no need to redeploy resources from or to any of these types of area. There was, for example, no need to concentrate the resources on the cities or the rural areas. Needs if they existed were specific and not related to type of area. The same is true of the geographical regions. The survey produced no evidence of regional differences and therefore no argument for redeployment of resources. There was no argument for more schools, more teachers or more instruction materials in one part of the country than in another. Educational priority areas such as those established in the 1970s could not simply be defined in terms of general types or in terms of geographical region. Much more specific information than that was needed and therefore much more focused intervention.

Another apparently negative piece of information but one still relevant to

policy, both national and regional, was the finding that pupils from smaller schools attained practically the same standards as those in larger schools. 'In particular, pupils from one-teacher schools reached the same standards as those attained by pupils in schools where there were more than six teachers' (SCRE, 1963, p189). Since the publication of the reports, we have experienced, in Scotland, in England and no doubt elsewhere, the closing of small one- and two-roomed rural schools. The evidence of the survey made it perfectly clear that such action was not justifiable on the basis of pupil achievement. The arguments for these changes which proceeded on a massive scale in the 1960s and 1970s has to be on the basis of cost or other social values.

The second survey produced more, and equally valuable, information for the formulation of policy. It showed that gains in achievement were greater in those areas where the local authority continued to make use of a battery of attainment tests on the completion of primary schooling. An obvious inference would be that the existence of a formal external assessment of this kind has beneficial effects upon the attainment of pupils. The information about library provision in primary schools has equally important implications. 'Pupils in schools with libraries of various types have made higher scores in the English test than pupils not having these facilities' (SCRE, 1968, p80). While, as noted above, the researchers are cautious, they do go on to conclude 'nevertheless the association shown between possession of a library and the high performance on English tests is suggestive' (p80).

The process of policy formation is one that is not easily unravelled but there is no evidence that even one small school was spared because of the findings of the research. Certainly I have not seen it cited during the debate that has taken place over the last ten years and which continues today. The arguments for closing small schools are predominantly economic, though the social development of the children is also mentioned and occasionally fears, apparently misplaced, are expressed about academic achievement. The protagonists of the schools usually advance community values and the deleterious effect of travelling on their side.

The impact of the finding on the use of attainment tests at the end of primary schooling is clearer. All authorities have now abolished them in spite of the evidence that their use was positively related to improvements in attainments. I have found no evidence that the provision of school libraries has been based on the findings of the scholastic surveys. In the present period of retrenchment I have seen no reference to the importance of maintaining the school libraries because of their anticipated effect on achievement in English. As far as I can see the recommendations which had relevance to national policy have been ignored. Why was this?

In a recent analysis by the Rand Corporation of the contribution of evaluation to policy, *Educational Evaluation in the Public Policy Setting*

(Pinkus, 1980), a number of points were made that relate to this issue. The authors point to timeliness, costs and values as important factors.

If the evidence from surveys is to help determine national policy then the information provided must speak to contemporary concerns. It is unlikely that accidental evidence gleaned in the process of a survey and simply recorded in technical reports will have any influence. Specific information from explicitly focused studies must be produced at the appropriate time.

The information from survey studies is partial and may therefore be misleading. Perhaps the administrators who abolished attainment testing in spite of the survey evidence were right. A notion of what primary education is meant to achieve is not adequately defined by formal tests. Any advantage accruing to schools from use by the authority of attainment tests may well be outweighed by other more negative effects on the curriculum of the schools. As the Rand Corporation report points out 'Studies that use a single outcome score to judge the relative value of programmes without regard to different programme goals or approaches are of little value. Large scale summative evaluations should be reconceptualized ...(to) present carefully justified judgements about the relationship of programmes to changes in educational treatments that may be affecting children' (p84).

As the authors of the 1953 report wisely point out 'an analysis of the effect of class size has yielded no clear conclusions. It appears that an investigation of this topic would require a specific design in which the accepted principles for organising classes would be altered for the purposes of the experiment' (p168). This is a conclusion which might well have been applied to other findings about school size, school libraries or the use of external examinations. Administrators were rightly sceptical of conclusions based exclusively on formal tests of arithmetic and English and which could not take into account a full range of contextual variables. More focused studies related to the effects of particular administrative arrangements are necessary to provide a balanced picture for the guidance of policy makers.

Teachers' planning

What contribution did the surveys make to the second set of objectives? That is, what help was available to teachers for planning the balance of pupils' work?

Teachers have two interests. The first is in the standards of their own pupils compared with those in other similar schools, as with wages our reference groups tend to be local and individual rather than national and general. It is a question of each teacher defining for himself what standards are appropriate in his circumstances, finding out whether his pupils are reaching those standards and taking the appropriate action.

The teacher's other interest is what he should teach and how he should teach

it. National surveys cannot help any individual teacher to decide what teaching scheme should be used next year, nor how it should be used and still less the balance of work for particular pupils.

The specific advice to teachers in the scholastic survey illustrates these limitations. Did many schools cease division by factors, or pay more attention to the layout of short division sums or provide persistent oral practice in accepted speech forms as a result of the publication of these findings? If they did, how many teachers would now think that was good advice? As with more general issues the specific recommendations relate to a particular perception of the purposes of school which is not now so widely held.

Even for those who do accept the assumptions of the authors, how would a particular teacher know whether the more attention which it is held should be paid nationally to the layout of a short division sum applies to his class? If he was already providing more attention than the average should he provide not more but perhaps less? Is it likely that those already giving 'considerable attention to the layout of short division sums' would feel strengthened in their conviction and provide even more? Would those not giving sufficient attention have overlooked this point in the recommendations of the report?

Findings from national surveys may or may not apply to any particular teacher and whether any teacher will take account of them will depend very much on their own values and their own perceptions of their current practice. In the case of the standardised notation for the recording of time of day the survey merely indicated variation in practice. The panel's choice of a particular form arose not from the survey but from their own general experiences. Information which will be relevant to specific questions cannot for the most part be satisfactorily obtained from a national survey.

A teacher's decision about the emphasis to be given to layout in teaching arithmetic is more likely to be based on his own experience of the situation around him than on any information that the nation as a whole did well or did badly in this respect on a general test. Individual teaching decisions are not made on the basis of general tests but on the basis of specific information which relates to the teacher's own objectives in the circumstances in which he is operating.

The general trend of the report is towards a greater standardization of curriculum and method. It is difficult to see how it could be otherwise. A few quotations will illustrate the point. 'Examination of the errors demonstrates forcibly the need for persistent oral practice of correct forms and usages...Although pencil and paper must be used for testing this kind of usage, it is not the best medium for teaching' (SCRE, 1963, p106). 'The use of the apostrophe...is not taught...the correct form should be shown and explained' (p107). 'It would appear desirable to include some printed Scots among the reading material for Scottish children' (p100).

The advice may be good or bad but the conclusion is clear. Either one has a national curriculum which includes those elements which the authors thought were important or one maintains the traditional British division of responsibility, placing trust in the professionalism of the teachers. The position of the authors of the report is akin to that of St Augustine when he prayed, 'Lord, make me chaste - but not just yet'. The report says that 'it would be folly to attempt to standardise curricula until it has been shown that one method is superior to others...If it were possible to determine a standard order to teaching...these contributions would be useful contributions to teaching' (p185) - but not just yet.

The authors of the survey wanted their cake and halfpenny as well. They wanted to have in effect a national curriculum but did not recommend so directly nor did they recommend any mechanism for establishing or enforcing it. Perhaps because they anticipated rightly that any such recommendation would meet overwhelming objection from the teachers' organisations.

In the case of the recommendations which were relevant to classroom practice, the authors of the report failed to grasp the nettle and draw the conclusion that was implicit in most of their recommendations i.e. there should be a national curriculum. If there was not to be a national curriculum then the recommendations were to individual teachers but as suggested above they were not a form which provided useful guidance to individual teachers.

Information for parents, employers and others

When we turn to the third issue, that of making employers, parents and others better informed, there are again a number of problems. Information about the current position and/or changes is a recurring concern of those involved in administration, of educational researchers and occasionally of those with a more general interest in the schools, such as employers and academics. There are occasional flurries of interest in national standards with headlines in the national press, but they are usually followed by a period of quiescence. The former British Prime Minister, Mr Callaghan, started 'a great debate' on standards in education. Little is heard of it now. Instead interest is focused on the effects of reduction in public expenditure.

The call for information about contemporary standards sounds reasonable enough but it is not at all clear what use this information has. There is for example, considerable American evidence that standards of candidates for the College Entrance Examinations have been dropping steadily in recent years but since nobody knows why, there is not much that can be done about it.

Recording changes as they occur is less obviously compelling on analysis. It seems self-evident that we should monitor standards over time as a sort of quality control but what use can be made of such general information? Such

findings are important because they correct false impressions. It is easy to believe that standards are falling. Two or three experiences with shop assistants who cannot perform simple arithmetic accurately and quickly would convince the casual observers that standards are low and indeed falling. Yet the data from the surveys I report and from later surveys indicate that virtually all school leavers have a high level of facility in rote arithmetic.

Similar important, if negative, findings were produced as part of a recent study of the primary schools (Scottish Education Department, 1980). These surveys demonstrated that the standards of achievement in the schools in arithmetic and reading were high and indeed in most aspects higher than they had ever been. This meant that the inspectors in their part of the report could go beyond the sterile arguments about falling standards in the basic skills to look at what primary education ought to be concerned with. When *Rising Standards in Scottish Primary Schools* (SCRE, 1968) was published ten years ago, however, it was not exactly a best seller. No-one seemed to want to know. Perhaps the problem lay in the title. Would a book entitled *What has Happened to Standards in Scottish Primary Schools* have sold better?

It is arguable that what parents and employers and others need is not more information of a general kind about standards but a better understanding of what it is that schools are setting out to achieve and how particular activities fit into these objectives. Employers need to know, as a basis for discussion with educational authorities, what arithmetic the schools are trying to teach and what communications skills are being taught. Parents need to know that apparently random play activities in Primary 1 or field studies in Secondary 4 are carefully thought out parts of an overall programme making a specified contribution to children's learning. They also need to be reassured that the schools their own children attend are providing the same opportunities as are available to others. Surveys of national standards will not inform them on either of these points.

There are circumstances when national surveys can be useful for national policy-making. These are mainly when the national conscience is agitated by a specific educational issue. If there is concern about standards then national surveys may play some useful role in providing empirical evidence. Even then as an article by the Secretary of the Scottish Confederation of British Industry (the national association of employers) demonstrates, there may be a tendency among the protagonists in the debate to question the survey evidence (Devereux, 1979).

Surveys may also have a publicity value in some circumstances. The publicity given to a series of surveys on reading standards contributed to the atmosphere which made the establishment of the Bullock Committee acceptable.

Dissemination

How did the Council expect to affect practice? The reports that they published were highly technical and were presumably addressed to the research community. Mulkay (1977), when drawing the familiar distinction between pure and applied research asserts that for pure research 'the audience for results consists of other researchers who are working upon the same or related problems and have judged the adequacy of the results by means of scientific criteria' (p95). That audience is interested in the extension of scientific knowledge. Where the findings are expected to 'have useful practical consequences' (p95), other criteria apply and other kinds of communication are appropriate.

The Scholastic Survey Committee made a conscious attempt to reach at least two of the audiences referred to above by means of an abridged report called *The Attainments of Scottish Ten-Year-Olds in English and Arithmetic* (SCRE, 1969). It was published 'in accordance with their policy of making research findings available in compact form to teachers, parents and others' (p2). This report consisted largely of the two tests and the technical material. It did have a chapter devoted to general results where the major findings about sex differences, different types of area, different size of class, different sizes of school and the significance of left-handedness were reported. However, the information of significance for teachers is buried in the analysis of the test scores. Even the briefer report explicitly designed for lay audiences seems to have been written with rather more than half an eye to the research community.

Caplan, Morrison & Stanbaugh (1975) outline three utilisation theories which seek to explain problems of communication between the social researcher and his audience. The 'knowledge-specific' theories try to explain lack of use of social science knowledge as a consequence of the nature of the information itself and the research techniques employed. The 'two communities' theory explains failure to use research in terms of the relationships of the researcher and the research system to the policy-maker and the policy-making system. Finally, the 'policy-maker constraints' theories argue that failure to use can best be understood from the stand-point of the constraints under which the policy-maker operates, for example, his need for concise information in a short period of time. In the case of the Scottish surveys all three sets of problems existed. The reports were addressed to the research community and not presented in a form which was likely to attract the attention of any one of the lay audiences to which the conclusions were presumably addressed. 'The research is focussed on understanding and fails to provide necessary action frame-work' (p. x).

While the Committee were anxious to draw attention to the practical implications of their findings, they seem to have no thought to 'key points where it will be most likely to be used' (p. xi), thus maintaining the barrier

between the two communities. Nor do the researchers seem to have taken into account the constraints on the policy-makers, the extent to which other factors must determine the decisions actually taken.

The reports, as distinct from the conclusions, do not seem to have been addressed to the relevant audiences and were hence likely to get lost in the theoretical literature rather than reaching those who were in a position to use them.

Conclusions

What can we learn from the Scottish experience? Our French colleagues have a reputation for pithy comments. You may know that when in 1918 Wilson produced his fourteen points, Clemenceau commented 'le bon Dieu was satisfied with ten'. I am afraid I can measure up to neither. I have only seven. First, if we wish to be listened to, at least in the short run, we must speak to policy-makers about the issues that concern them when they concern them. Second, we must recognize that our contribution to the discussion is a partial contribution. There are other considerations, economic, social, political, which may over-ride our findings, no matter how conclusive we think them. Third, what is sought is usually knowledge of specifics which is relevant to particular local circumstances. Studies on the grade scale may be interesting to researchers and utterly uninformative for policy-makers. What they require are focussed studies which provide information about particular issues. Fourth, it may be necessary for us to sacrifice some of our academic purity to provide information which will be of help for policy formation. Fifth, national tests will inevitably have some curriculum backwash and will involve pressure towards a centrally determined curriculum no matter what we may wish. Sixth, if we wish to influence classroom practice our findings must, in Eaker & Huffman's phrase, be not only statistically tested but also classroom tested (1980). Finally, administrators, politicians, teachers, parents and employers will not take us as seriously as we take ourselves.

2

On Intelligence

This essay was written as the introduction to the published edition of the papers presented at the Toronto Symposium On Intelligence, held in 1969. The conference papers which are referred to throughout Dr Dockrell's introduction may be found in ***On Intelligence: Contemporary Theories and Educational Implications: a Symposium, Toronto, 1969****, ed Dockrell (Ontario Institute for Studies in Education, 1970)*

Intelligence has been a concept of great significance in psychological theory and educational practice. However, challenge to this concept on both sides of the Atlantic has resulted in widespread re-examination of principles and practices which were previously accepted. This symposium was organised, therefore, to further the examination of basic theory and educational practice in the light of recent research.

The theoretical importance of the concept of intelligence for psychology hardly needs to be demonstrated. While it is true that the predominant role of the concept and investigations into it, which were a feature of the psychological journals of the second and third decades of this century, no longer exists, nevertheless, prominent psychologists continue to produce books and articles on this topic and there are journals devoted primarily to publishing research in this field. Intelligence remains a major concern of psychology. The educational importance of the concept can be seen both in research and in practice. A casual survey of the research journals in education shows that the concept of intelligence is used as an experimental or control variable in well over fifty per cent of the studies reported. Critical examinations of the concept are few but its value is assumed in most educational research.

Application to educational practice varies, perhaps with ideology. In England, the tripartite system of secondary education was justified on the grounds of differences in intelligence, but has been criticised in part for its inefficiency in sorting out the bright from the less able. In Canada, where there has been little attempt to provide a rationale for the educational system, the influence of the concept of intelligence is most apparent in the provision for children typically classed as 'educable mentally handicapped'. Occasionally, provision is made also for the other end of the spectrum, the gifted. The position in the United States has been broadly similar to that of Canada. There, provision for children of different levels of success in school learning has usually been made on an *ad hoc* basis within the schools. With rare exceptions separate special provision is made only for extremely poor learners, classified, as in Canada, as educable

mentally handicapped. At the tertiary level of education, however, even in the United States, colleges and universities typically make use of 'aptitude' tests which are taken to measure something other than the knowledge and skills explicitly taught in the schools. While the word intelligence is avoided, the concept is not.

The whole notion of intelligence, both as a theoretical concept and as a guide to educational practice, has been criticised from the beginning (Watson, 1930). Indeed, the relative importance of cognitive structures and environmental experiences had been a source of dispute in education and in the philosophical antecedents of psychology long before the modern formulations of the concept of intelligence (Priestley, 1774). In recent years, the attacks on the theoretical basis of the notion of intelligence have come from the behaviourists, both in Russia and the United States (Skinner, 1961; Luria, 1961). The questioning of the utility of the notion for education has come primarily from sociologists (Halsey, 1958). Yet, much educational thinking retains an ability variable. A simple model of learning used in the international study of achievement in mathematics (Husen, 1967) has three components: previous knowledge, motivation and intelligence. Learning is a function of the interaction of these three variables. The major task for this symposium was to examine the usefulness of the third component of the model.

Much of the dispute, both in classical learning theory and in education, has turned on the relative importance of each of these variables. The relevance of the other two variables in specific learning situations is not disputed by the participants in this symposium. Indeed, the senior contributor, Burt, has elsewhere reported investigations into the importance of motivation (Burt, 1961). The sole question at issue is whether the concept of intelligence as a factor in learning, which is independent both of previous knowledge and of motivation, is theoretically fruitful and practically helpful. Does the notion of intelligence still help forward our thinking about learning as it appeared to do in the first part of this century, and does it help in our planning for teaching and learning?

In psychological theory, attempts to accelerate the acquisition of conservation as defined in Piaget's work (Sullivan, 1967) are frequently intended to show that conservation is a function of previously acquired knowledge. Similarly, Ausubel and his associates (Ausubel, 1967, Ausubel & Fitzgerald, 1962) have tried to show that what appear to be differences in motivation and ability are largely differences in previously acquired knowledge. Traditional studies of intelligence have attempted to control this previous learning variable and to demonstrate systematic differences in ability by studies of children raised outside their own families and by studies of separated twins (Burt, 1966).

Much of the theoretical dispute about the significance of intelligence as an

independent variable has been related to racial and social class differences. Some studies have emphasised either the previous learning variable (Davis, 1948; Hess & Shipman, 1965) or motivation (Haggard, 1954; Zigler & Butterfield, 1968) though some have stressed genetically determined differences in intelligence (Burt & Howard, 1957; Jensen, 1969).

Two examples from educational practice will suffice to show the concern with the significance of the three components of the learning model. The initial scientific impetus for the Headstart Programme in the United States came largely from studies involving intelligence (Hunt, 1961), but many of the programmes have emphasised the importance of previously acquired knowledge (Bereiter & Engelmann, 1966) or motivation (Zigler & Butterfield, 1968). In Britain there has also been an increased stress on motivation as a factor in ultimate educational attainment where previously the emphasis was on intelligence. Contrast, for example, the emphasis in the Plowden Report (Ministry of Education, 1965) on parental attitude with the concern with types of ability in the Haddow Report (Board of Education, 1926) and the Spens Report (Board of Education, 1938).

The crucial unresolved question before the symposium was whether the intelligence variable should be retained in the model, and if it should, what was its relative importance compared with each of these other two variables? In view of the wide range of human activities, where ability independent of previous experience and motivation seems to be important, the hypothesis that there is an ability component in human learning seemed plausible and worth the consideration of a symposium.

A basic problem for those who wish to investigate the ability component in the learning model is the extent to which intelligence is thought of as a convenient abstract generality like beauty or honesty, or as the behavioural correlate of some characteristic of the brain, possibly neurological, possibly biochemical. Koch has recently attacked psychologists who come 'to the conclusion that man is a cockroach, rat or dog... a telephone exchange, a servo-mechanism, a digital computer, a reward-seeking vector, a hyphen within a S-R process, a stimulation maximiser, a food, sex, or libido energy converter, a utility maximising game player, a status seeker, a mutual ego titillator, a mutual emotional (or actual) masturbator' (Koch, 1969, p14). Yet, each of these formulations has contributed something to our study of man. Koch is pointing out the risk of being carried away by a useful analogy and therefore seeing man as no more than a cockroach, rat, or servo-mechanism.

Oppenheimer, in an address to the American Psychological Association (Oppenheimer, 1955) argues for the inevitability of analogy in scientific thinking 'the conservation of scientific enquiry is not an arbitrary thing; it is the freight with which we operate; it is the only equipment that we have. We cannot learn to be surprised or astonished at something unless we have a view of how

it ought to be; and that view is almost certainly an analogy (pp129-30).' But he goes on to point out the dangers of analogy 'especially when we compare subjects in which the ideas of coding, of the transfer of information, or ideas of purpose, are inherent and natural, with subjects in which these are not inherent and natural (for then) formal analogies have to be taken with very great caution' (p134).

Thomson (1950) made this same point about the study of intelligence. He insisted that it is important that 'G (general intelligence) is interpreted as a mathematical entity only, and judgement is suspended as to whether it is anything more than that' (p240). He went on to examine the concept of intelligence as 'mental energy'. He pointed out that mental energy could not convey exactly the same meaning as physical energy, but he continued:

> if 'mental energy' does not mean physical energy at all, but is only a term coined by analogy to indicate that the mental phenomena take place 'as if' there were such a thing as mental energy, these objections largely disappear. Even in physical or biological science, the things which are discussed and which appear to have very real existence to scientists, such as 'energy', 'electron', 'neutron', 'gene', are recognised by the really capable experimenters as being only manners of speech, easy ways of putting into comparatively concrete terms what are really very abstract ideas. With the bulk of those studying science there exists always the danger that this may be taken too literally, but this danger does not justify us in ceasing to use such terms... the danger of 'reifying' such terms or such factors as GV, etc., is however, very great...(p251).

The different concepts of intelligence held by the participants in this symposium minimise the danger of accepting any one point of view about intelligence as correct. There remains the danger of unconsciously reifying the concept of intelligence and treating it as though it were an entity and not merely 'a convenient manner of speech'.

This problem is greatest, as Thomson says, in studies which make use of factor analysis. It is important to note that this mathematical technique does not speak to the issue of the validity of a particular concept of intelligence. All it does is make use of one of a particular group of mathematical procedures to arrive at a simpler set of hypothetical tests or factors, taken to underlie performance on a wider range of more complex real tests. Guilford (1967) makes the familiar distinction between a mathematical factor and a psychological factor. A mathematical factor is obtained by administering a number of tests to a group of subjects, correlating them, and following conventional mathematical procedures. A psychological factor, however, is a mathematical factor which is also 'conceived to be an underlying latent variable along which individuals differ, just as they differ along a test scale' (p41). But as Thomson (1950) pointed out, we cannot automatically infer a psychological factor from a mathematical factor 'it is then for the psychologist to say, from a

consideration of the ... tests which define it, what name this factor shall bear and what its psychological description is. The psychologist may think, after studying the tests, that they do not seem to him to have anything in common, or anything worth naming and treating as a factor. That is for him to say' (p226). The mere existence of a mathematical factor does not speak to its psychological utility.

A decision about the probable psychological utility of a factor does not end consideration about its status. There remains the danger of treating these 'really very abstract ideas' as realities. Vernon (1950) asserts, 'factors should be regarded primarily as categories for classifying mental or behavioural performances, rather than entities in the mind or nervous system' (p8). Burt, however, allows factors to have either status, as components of a test battery or factors of the mind. The danger in this case is in assuming that because a factor has practical utility as a component of a test battery that it is therefore a factor of the mind.

Take, for example, the contrast between Burt's and Merrifield's papers in this symposium. Burt defines intelligence as 'innate, general, cognitive ability'. Merrifield uses Guilford's model and talks of 120 factors. Does the mind consist of one broad general ability with other smaller less important groups of abilities, or of 120 independent abilities which may be summated in various ways for various purposes? If factors are thought of as convenient generalisations, the question is not whether there is one ability or many, but which model is useful in a particular context or for a particular purpose. If the question is a broad question, 'Am I likely to do well in a general programme involving arts and science subjects or not?' Burt's model seems most appropriate. If, on the other hand, the question is very specific, 'Am I likely to do well as an historian primarily concerned with bibliographic research or not?', the model Merrified adopts may be more useful. The question becomes not is Burt's or Guilford's model right, but is it appropriate. Does it help me to think fruitfully about a problem that is puzzling me, if I use Burt's way or Guilford's? Does it help me to make decisions about a particular question of educational practice if I use Burt's way or Guilford's? If we accept Vernon's position and view factors as categories for classifying mental or behavioural performances, the choice of categories would depend on the problem to be solved, or the question to be answered.

In Evans's paper, for example, what is the status of his factors? He chose certain tests, administered them to a sample of a defined population, and submitted them to specified mathematical procedures. There emerged certain factors which he could either accept as the basis for psychological speculation or reject as meaningless. He argues that his factor pattern is psychologically meaningful. Further, he seems to think of at least some of his factors as having a physiological basis. He refers to one of his factors as 'innate cognitive capacity'.

The factor that is of most interest to him, however, is 'Problem Performance'. He relates his factorial findings to a number of studies from other fields of psychology and, on the basis of theory, postulates a specific significance for this factor. This argument, however, speaks only to the psychological utility of the factor, not to its probable status. He hypothesises that this variable will emerge in specified circumstances. Is there then a physiological basis for the factor? Is it merely a way of classifying performance, useful in certain circumstances, or is it conceived of as in some sense a stable entity that can be developed by appropriate training? Is he measuring a set of related tasks which may be conveniently grouped together or some manipulable entity; a factor of the mind or a component of a test battery?

The same questions about the status of factors can be asked of Vernon and Jensen. Jensen had discussed elsewhere (Jensen, 1969) the basis for his assumption, that the differences in his level I (rote learning ability) and level II (problem-solving) are genetically determined. If he could indeed demonstrate that his factors correspond to some genetically transmitted physical basis, his model would be a criterion, a touchstone, for other psychological theories. However, an alternative position stressing the role of learning seems equally plausible to many psychologists (Hunt, 1969; Kagan, 1969). Rote learning ability (level I) may, as Jensen argues, be a necessary but not sufficient condition for the emergence of problem-solving (level II). The additional necessary condition, though, may not be an independent genetically determined ability, but the right kind of environmental experiences. The problem-solving strategies which Jensen discusses - for example, grouping items on a logical basis in order to remember them more easily - may be taught in one environment, but not another. Recent research by Kagan (1968) suggests how this might come about. Similarly, Guinagh's (1969) findings that children high in rote learning ability from low socio-economic status backgrounds could improve in problem-solving after a specific teaching programme, supports the hypothesis that the right kind of environmental experiences might, indeed, be the relevant variable.

The evidence that a restricted environment has its greatest effect on animals who are bright (Cooper & Zubeck, 1952) fits in with the environmental argument. One would expect then that children from an environment which did not encourage the development of problem-solving strategies would make very low scores on tests of this ability, even though they were relatively high in the basic rote learning skills. Children from an environment which facilitated the development of problem-solving skills would, however, score well on a test of such skills, but only if they had the necessary basic rote learning ability. One would, then, on the environmental hypothesis, hypothesise the kind of distributions that Jensen proposes on a genetic basis in his Figures 8 and 9. The argument for an independent physiological base for these two factors is, then, speculative and disputed.

We must therefore apply to Jensen's factors the same two questions that have been raised about other factors. What is the psychological plausibility of these factors, and what is their assumed status, test component or factor of the mind? Jensen reviews extensively in his paper the degree to which his formulation corresponds to other research findings. He makes a persuasive case for accepting the probable psychological utility of his factors. The answer to the second question, their presumed status, is less clear.

If the two abilities are transmitted genetically, presumably they have some physiological base and definable objective existence. Yet, Jensen asserts 'level I and level II are ways of conceptualising two broad sources of variance'. Are they merely useful constructs and not realities which are criteria for other models? If so, we may go on to examine their usefulness as a basis for action.

In his discussion of the status of his factors, Jensen comments that 'level I and level II ... may be further fractionated by factor analysis, that is, there are alternative ways of breaking down these test scores into other kinds of components'. His model then is one of several possible equally acceptable models. The question is whether his model is more useful than the alternatives in suggesting ways of tackling educational problems.

One of the most interesting sections of Jensen's paper is his discussion of the relevance of his theory to education and his suggestion for developing procedures, which would logically follow from his theory. There are, however, a number of problems with his approach.

As Jensen himself points out, children with different backgrounds use different patterns of abilities to solve the same problem. Is Jensen's model of intelligence subtle enough to detect all the differences in social class patterns of abilities that are relevant to academic success? It is possible that a model like Jensen's which consists simply of two broad abilities might not pick up differences between social class groups which are important for success in school. An alternative model like Guilford's, which breaks down ability into more precise components, might be more sensitive to the abilities of children who do not now succeed in school but could with appropriate teaching. Though Jensen is careful to point out that he is not advocating any over-simplified rote learning instructional programme, nonetheless his theory as such does not provide the educator with any more subtle or sensitive basis for detecting the abilities of children raised in less stimulating environments.

A further problem that might arise in trying to apply Jensen's concepts to educational practice is not unlike that faced by the British secondary school system in the 1940s and 50s. The three different types of secondary school were meant to cater to three types of minds. The variables, however, were continuous and not dichotomous, that is most children did not fall neatly into the three categories, but fell somewhere in between. Similarly, it seems likely

that, in Jensen's terms, there would be children high in both abilities, children high in rote learning but low in problem-solving, and children low in both; but it is also likely that most children would fall somewhere in the middle and be hard to categorise. Educational programmes based on a simplified model and designed for pure types would probably have very limited application.

Certainly, it might be worth trying to develop educational procedures on the basis of Jensen's theory, but their usefulness remains to be demonstrated.

In assessing the contributions to this symposium then, it will be important to bear in mind Thomson's warning. Many of the concepts of science are 'only manners of speech' and it is dangerous to take analogies literally. This is particularly true of psychology where the alternatives have often appeared to be either a sterile concentration on specific behaviours or heady generalisations, both very difficult to apply to practical situations.

Burt admirably set the stage for the symposium with a survey of the history of the concept of intelligence and its relevance to contemporary issues. The contributions of Evans, Jensen and Vernon suggest that intelligence as a theory is still a fruitful basis for thinking about human learning. Tuddenham showed that conventional psychometric techniques are a way of operationalising theoretical thinking like Piaget's, derived from an entirely different frame of reference. As for educational practice, Jensen is proposing a specific approach to an important educational question, how best to educate a large segment of those who do not succeed in school. Vernon provides a theoretical basis for educational procedures for students from cultures radically different from the ones where current educational values and practices were developed. Merrifield shows how the most recent major development in theorising about intelligence may be applied to educational practice. These contributions to the symposium suggest that intelligence as a concept is alive and well, providing fresh insights for theoretical problems and making new contributions to the practice of education.

The Warburton paper is of particular importance to the symposium, though its content is of interest primarily to psychologists in the schools, for it was the knowledge that Warburton and his colleagues in Manchester were developing a new individual intelligence scale that led us in Toronto to think again about intelligence and to call this symposium. It is with gratitude and respect, therefore, that this report of the symposium is dedicated to his memory.

3

Assessment in the Classroom

A paper delivered in San Francisco in 1986 at the annual conference of the American Educational Research Association (AERA).

Introduction

Can teachers prepare criterion-referenced tests which relate to their own classroom teaching? Can they use them to highlight the needs of individual pupils and to indicate what problems there are in the curriculum or in the instruction? What help will teachers need to make and use such tests?

These were the questions we began with when we launched our programme of studies. We began with a small number of case studies working primarily with children of junior high school age, roughly the equivalent of grades 8 and 9. We decided to investigate three areas of the curriculum: one, where the teaching was modular; one, where learning was taken to be linear; and one, where the emphasis was primarily on the acquisition of skills. To exemplify the first area we chose geography; to exemplify the second we chose foreign languages; and to exemplify the third area, we chose what in Britain are called Technical Studies and I think in the US are called 'Shop'.

The Tests

In the geography syllabus there was a unit on the environment. The teachers prepared a test composed of items which related to the six concepts they were trying to develop. They used the test for two purposes, one related to the performance of individual pupils. Instead of simply giving a total score and saying to some pupils 'you were very good'; to others 'you are middling good'; and to others 'you are very bad'; they were able to say to even the highest scoring pupil 'you were very good, but you do seem to be having problems with this aspect of the unit'; to those in the middle, they were able to say 'your problems are related to these particular aspects of the unit'; and to those pupils who made the lowest score, they were able to say 'you did well on this aspect of the unit, but you are having difficulty with the others'.

The approach of the geography teachers was to think of their curriculum in terms of core and extensions, if you like the necessary and the nice to have. Everybody should master the core. After the test had been administered therefore, the pupils were given the appropriate remedial work and where they completed the remedial work before the end of the time allocated to the unit, extension work in the same topic.

Table 3.1 shows the results. Failing more than one item was taken to indicate failure to master the concept. The teachers were not able to wave a magic wand and turn all the ducks into swans. First time round 26 pupils in the class for which the tests are shown in Table 5.1 failed to meet the requisite standard in 54 areas among them. After the remedial work this number was reduced to 26 areas.

Table 3.1: Feedback from a Diagnostic Test on the Environment – the section scores of a class on the 'Environment' test

Pupil	Conservation	Vandalism	Dereliction	Man-affected Environment	Natural Environment	Pollution	Total
A	6	2	3	3	(2)	5	21
B	5	2	3	4	(1)•	4	19
C	(4)•	(1)•	(2)•	3	3	4	17
D	(3)	(1)•	3	3	(1)	5	16
E	5	2	3	3	(2)•	4	19
F	5	2	4	(2)	(2)•	4	19
G	(4)•	2	3	(2)	(2)	5	18
H	6	2	4	3	(2)•	5	22
I	(4)	3	3	3	(2)	5	20
J	5	2	3	(2)•	3	5	20
K	5	2	3	(2)•	3	5	20
L	(4)•	3	3	(1)	3	5	19
M	5	(1)•	4	4	4	5	23
N	5	2	4	3	(2)•	5	21
O	(4)•	2	4	3	4	5	22
P	(3)•	2	4	(2)	(2)	(3)•	16
Q	5	2	4	3	(2)•	5	21
R	6	2	4	3	(2)•	5	22
S	5	2	3	(1)•	3	4	18
T	5	2	3	3	(2)	4	19
U	(4)	3	3	(2)	3	4	19
V	(2)	3	(2)•	(2)	3	4	16
W	(4)	2	3	3	(1)	4	17
X	(0)	2	(1)	(2)	(2)•	(3)•	10
Y	(3)•	2	(2)	(1)	(1)	(2)•	11
Z	(4)	2	(1)•	(1)	(2)•	4	14
Number of items in section	6	3	4	4	4	5	
Pass score (arbitrary criterion)	5	2	3	3	3	4	

(O) Students failing to attain pass score.
• Fail modified to pass on post remedial test.

The teachers could use this information not only for individual diagnosis and remediation, but to examine the whole curriculum. Some concepts were clearly more difficult than others. After the completion of the remedial work all pupils in this group had mastered the notion of vandalism and pollution but many had still not mastered the concepts of conservation and natural environment. Were these concepts appropriate at this stage? If they were should the curriculum be revised to take account of the difficulty level of the material?

Table 3.2: Percentage of Students Attaining given Concepts in Geography Settlement Unit

class	*optimum site*	*service field*	*spatial differences in cities*	*uniformity in cities*
A	7	89	64	36
B	7	100	81	40
C	29	95	62	38
D	52	100	85	48
E	59	96	93	41
G	11	100	89	19
H	25	92	71	17
L	0	96	78	44
M	28	100	76	40
N	33	100	63	38
R	77	92	73	23
V	8	92	73	27

Attainment of concept is at least ⅔rds of the items correct for each domain.

There are questions to be asked too about instruction. Table 3.2 shows the results for another unit broken down by concept and class. The classes were all what we call mixed ability groups, that is they were grouped heterogeneously. Yet there were wide differences in the numbers achieving the required level in different classes as can be seen from Table 3.2. Questions, therefore, had to be asked about the instruction in particular classes. It is not a question of generally bad teaching, but apparently of the different emphasis given by specific teachers to particular concepts.

An example of the type of tests which were developed jointly by the teachers and the researchers in foreign languages was one for the use of the dative pronoun in German. This concept is particularly difficult for British students. The teachers were asked to draw on their experience of teaching this aspect of

their course to hypothesise the sorts of common errors that might be expected. Thus it was hypothesised that in a situation where they should be using the dative plural pupils tended to use the masculine dative singular (error 1), the feminine dative singular (error 2), or the dative second person plural (error 3). An extract from the test is shown in Table 3.3.

Table 3.3: An Extract from the Test for use of the Dative Plural in German and the Item Rationale for the Whole Test

Class III Pronouns Test

If you replaced the underlined words by a pronoun, the correct answer would be A, B, C, D. Put a tick in the appropriate box.

1 Ich spiele mit den Kindern

a Ich spiele mit ihm
b Ich spiele mit ihr
c Ich spiele mit ihnen
d Ich spiele mit Ihnen

2 Wir gehen mit den Mädchen spazieren

a Wir gehen mit ihnen spazieren
b Wir gehen mit ihm spazieren
c Wir gehen mit Ihnen spazieren
d Wir gehen mit ihr spazieren

3 Die Klasse sitzt vor der Lehrerin

a Die Klasse sitzt ihr
b Die Klasse sitzt vor Ihnen
c Die Klasse sitzt vor ihm
d Die Klasse sitzt vor ihnen

4 Der Junge ist bei seinen Schwestern

a Der Junge ist bei ihm
b Der Junge ist bei Ihnen
c Der Junge ist bei ihr
d Der Junge ist bei ihnen

5 Der Mann spricht mit den Frauen

a Der Mann spricht mit ihr
b Der Mann spricht mit ihm
c Der Mann spricht mit ihnen
d Der Mann spricht mit Ihnen

ITEM RATIONALE

	OPTIONS			
Item	*a*	*b*	*c*	*d*
1	I	II	•	III
2	•	I	III	II
3	Singular			
4	I	III	II	•
5	II	I	•	III
6	III	•	I	II
7	Singular			
8	II	I	•	III
9	Singular			
10	•	III	II	I
11	III	I	•	II
12	•	III	II	I

KEY
I — Error I
II — Error II
III — Error III
• — Correct

Approximately two thirds of the way through their first year of German the pupils were given this test. They were given the results but the scripts were not returned to them. During the following week the scores on the test were used

as a basis for remedial teaching. The scores on the first and second administration of the test shown in Table 3.4.

Table 3.4: Summary of Pupil Responses to a Test for use of the Dative Plural in German

		TEST 1				TEST 2			
		Correct	Error I	Error II	Error III	Correct	Error I	Error II	Error III
Pupil	A	6	0	0	3	7	0	0	2
	B	6	0	0	3	9	0	0	0
	C	1	3	3	2	5	1	3	0
	D	4	2	2	1	6	0	2	1
	E	5	0	0	4	5	0	0	4
	F	4	3	2	0	6	1	1	1
	G	5	0	3	1	9	0	0	0
	H	9	0	0	0	8	0	0	1
	I	5	1	1	2	6	0	1	2
	J	3	0	4	2	5	0	3	1
	K	5	0	0	4	9	0	1	0
	L	9	0	0	0	7	0	2	0
	M	9	0	0	0	8	0	1	0
	N	3	0	3	3	7	0	1	1
	O	1	4	1	3	8	1	0	0
	P	9	0	0	0	9	0	0	0
	Q	9	0	0	0	8	1	0	0
	R	5	2	2	0	4	1	4	0
	S	9	0	0	0	9	0	0	0
	T	2	3	4	0	6	0	0	3
	U	9	0	0	0	8	1	0	0

Notes: 1. *Only the nine items requiring a dative plural response are included in this analysis.*

2. *Only pupils who take the test twice are included in this analysis.*

For our purposes we have assumed that more than one error indicated a significant problem in a particular area. Clearly some pupils had mastered the concept, others were showing a whole range of errors, but other pupils where showing specific errors. There were twenty three specific errors the first time round. After the remediation nobody made more than one type 1 error, five pupils made more than one type 2 error, and four made more than one type 3 error: 10 errors in all. Not only could the tests be made, they could be used to improve the performance of individual pupils.

An example of a test in technical studies, was a template which pupils could use to see for themselves whether their work lay within the acceptable tolerance levels or not: for example, a gauge was made to measure wooden pegs which where being made to use in a board for playing noughts and crosses.

Many of our curricula have an affective dimension: for example, one geography unit had as one of its purposes to increase pupils' sympathy towards the people of Third World countries. They used a test where each pupil was told that they were to imagine that their class had collected £20 to donate to a charity, that the money was to be given out in £5 units, and they were to state their choice of charities from the list provided. The test included cancer research in Britain, new sports equipment for their own school, as well as new health clinics for poor cities, famine relief and so on. The test was given before, and again after, the unit to measure change which took place as a result of the teaching. The purpose of tests of this kind was not to see which of the pupils in the classes were budding Geldofs, but to assess the affect of the curriculum as a whole.

In all cases it was possible to prepare tests which related to the curriculum as it was being realised in particular schools, and which picked out areas of difficulty for individual pupils, and/or could be used to highlight curricular or instructional issues.

The teachers' needs

We began with what proved to be a very naive assumption: that teachers knew clearly what it was they were setting out to accomplish, and that our task would be to provide them with some skills in test construction and some technical back-up. It soon became clear that very few teachers thought in terms of what it was they were trying to accomplish. It was often implicit but it needed a lot of teasing out. The type of statement the teachers made to us initially was most frequently a list of course content, with little or no indication of what was expected of pupils. In most cases teachers were concerned solely with recall of information. There was very little attention to the acquisition of concepts or skills.

The range of testing instruments which our teachers initially proposed was very limited. For example, in geography an objective of one part of the curriculum was to make pupils 'more aware of pollution'. How do you know when a youngster has become aware of pollution? How can you tell? The first response was to say, we'll get them to write an essay on the environment, but the very setting of the essay begged the question. Some more indirect methods where called for. In the end, the teachers came up with a series of photographs which covered a number of aspects of the unit but which also included an example of pollution, for example a photograph of a residential estate with a factory chimney in one corner puffing smoke, or a photograph of an attractive

valley with a stream running through it that had rubbish dumped into it. The children were then asked to list four features of these photographs. If in each case they included the pollution in their list of four features, they were taken to be aware of pollution.

The teachers did also require a great deal of help in test construction, even those teachers who had some course work in test construction were not familiar with the techniques of empirical and logical review appropriate for criterion-referenced tests. The materials suggested by the schools needed a great deal of rethinking and revision before they could be used in the schools.

The preparation of tests of this kind, for even a single unit, placed heavy demands on the staff of one school. We needed a co-operative effort which involved a number of schools preparing different sets of materials for common use. In Scotland this is possible. We do not have any mandated curriculum, schools are, in theory, free to devise their own. In practice, however, there is a great deal of commonality which is brought about by having externally-set, curriculum-based examinations at the end of grade 11 and grade 12. These examinations are passed by over three-quarters of the age group. There is too a National Curriculum Development Service which prepares materials which are used in most, but not all, schools. When we carried out a survey of the curriculum being used in mathematics, for example, we found that virtually all Scottish schools were using the materials prepared by the Scottish Mathematics Group. Uniformity varies from subject to subject, but it is much greater than one would find in, for example, England.

In some subjects at least, it was possible to prepare assessment instruments which related to what was substantially a common curriculum, and we have done so in geography and in technical studies. Shortly after we began our work, a new curriculum was being developed in foreign languages and the development team prepared their own formative test as an integral part of their materials.

The use of the tests

It was possible to prepare these materials, but how would they be used by teachers? There are significant variations in the way in which an apparently common curriculum is taught. Teachers select from the materials available to them, and put their own emphasis on particular aspects of the curriculum. What we prepared and made available, therefore, was what we called 'a resource'. For each unit of the curriculum we tried to have more material than any individual teacher could use. What we suggested was that the teachers should specify their own intended learning outcomes and select from the resource those items which related to their personal objectives; and that if they had objectives which were not covered satisfactorily by the resource, they could use our materials as a model.

Where there was no common curriculum, for instance in home economics, we did prepare a set of material which was intended partly as a resource but mainly as a model for teachers to follow in the construction of their own assessment materials (Black, 1983). We followed the classic British approach of selecting what we took to be good practice covering the full range of the curriculum.

The impact of the approach on learning and teaching

The experimental schools

We have attempted to evaluate the impact of our work in a number of ways. We interviewed pupils; we questioned teachers both in interviews and by questionnaire; we observed teachers' practice in their classrooms; and we assessed pupils' learning.

Pupils who had extended experience of this approach had a generally more positive attitude to assessment. Assessment was not seen as a weapon to be used by teachers against them, or as a means of control, but as a means of helping them to learn. They particularly appreciated getting feedback on their problems and the additional work that was given to help them overcome their difficulties.

The teachers were virtually unanimous in seeing the benefits of the approach. They thought, for example, that by making pupils aware of their particular problems, and of course their strengths, it made them more willing to seek help. As for themselves, the teachers reported that it increased their own motivation and, because they were aware of pupils' problems, they could organise their teaching more effectively.

Our classroom observation studies did show a substantial difference in practice between classes where the diagnostic approach was being used and those where it was not. As Table 3.5 shows, the lessons where our approach was being used tended to be more pupil-centred and individualised in their activities, and consequently placed greater demand for management skills on the teachers. While these lessons had more work-related pupil discussion, they also contained more disruptive non-lesson activities. This disruption generally took the form of chatter and did not represent a break down in classroom discipline.

I have referred earlier to the impact on pupil learning. More pupils attained what the teachers included in the core. Those who did not attain all the core, attained more of it: pupils moving to successive stages, therefore, had a better basis for later work. Successful pupils were stretched by being given extension work (and that is important in our context). And, finally, the assessments induced a more positive attitude to learning in pupils.

Table 3.5: Percentage of Observed Time Spent on Each of the Activities for a Selection of Groupings

	geography				*technical education*			
	all departments		non-individualised		individualised		all departments	
	NDA lessons	DA lessons	NDA lessons	DA lessons	NDA lessons	DA lessons	NDA lessons	DA lessons
1 Teacher lectures	19.7	7.5	16.8	15.2	22.2	2.8	12.0	10.1
2 Teacher instruction	9.3	15.4	13.5	11.8	5.6	17.6	7.9	7.3
3 Teacher-led questioning	20.5	11.0	19.8	17.0	21.1	7.3	11.0	15.8
4 Teacher management	7.3	6.7	6.9	7.0	7.6	6.5	7.5	9.8
5 Teacher authority	4.2	3.7	4.3	4.1	4.1	3.5	2.1	2.2
6 Pupil-led questioning	7.8	8.7	7.3	10.9	8.2	9.0	13.0	9.7
7 Pupil discussion	4.7	10.7	3.0	1.3	6.1	16.5	10.0	4.6
8 Teacher-centred work	16.6	7.0	15.9	15.5	17.2	1.9	19.0	21.6
9 Individual work	-	14.9	-	-	-	23.9	-	-
10 Co-operative work	-	-	-	-	-	-	1.8	3.1
11 Disruption	7.0	10.6	8.9	14.6	5.3	8.1	9.2	7.8
12 Non-lesson activities	3.1	2.8	3.6	2.6	2.6	2.9	6.5	4.1

NDA non-diagnostic assessment DA diagnostic assessment

We investigated a number of other issues which are, perhaps, more closely related to our circumstances than to yours. Our teachers, as I said above, were not used to defining the outcomes of their teaching in the way that we required. Most of them came to accept the approach and reported that it made them critical of their own previous practice and of the materials that they were using. It required them to be more careful in the preparation of their lessons. They reported too that the clear evidence of success in their pupils gave them a sense of achievement.

There were, of course, some difficulties. Teachers listed far more outcomes than they could possibly achieve. There was some difficulty in distinguishing the essential core outcomes and there was a tendency to accept, uncritically, the lists provided by the researchers.

Other schools and classes

That is what happened in the experimental schools which we were working with directly. What happened elsewhere? We have not carried out a formal survey of change in practice in Scottish schools over the period of our work. We are dependent on indirect measures. The most concrete is the take-up of our materials. There are about 450 schools in Scotland catering to our age group. We have sold 398 copies of our geography materials (Black & Goring, 1983), the great bulk of them in Scotland. Also, 135 schools have requested permission to copy our materials and adapt them for their own use as we said that they should. The home economics materials (Black, 1983) have been bought by 486 schools, approximately half of them in Scotland.

We are aware that our approach has been adopted by colleagues working in a number of different disciplines. As I stated above, when a new set of modern language materials were being developed they included a programme like our own. We have been invited, too, to help a number of schools which were adopting our approach across the whole curriculum. Most significantly from our point of view, the local education authorities jointly funded a unit to develop and extend our work.

Conclusion

It seems to us that we have demonstrated first, that our approach is feasible; second, that it is acceptable to teachers; and finally, that it has desirable effects in the schools.

4

Reporting Assessments of Pupils' Attitudes and Personality

Introduction

It has been argued that British secondary education has been dominated for over a century by the public examination system (Dockrell, 1985). It is not merely that the official school leaving certificate consists of reports of results in these external examinations but that they dominate the schools, for example, by determining the curriculum, school leaving reports (even for those pupils who do not pass public examinations) and assessment and reporting throughout the secondary school.

In recent years there has been a concern about all these consequences of public examinations including recognition of the need for more comprehensive reporting procedures such as records of achievement or pupil profiles. Both of these seek to encompass a wider range of attainments and characteristics than can be covered by public examination. There have been many local developments in the last decade (Dockrell & Broadfoot, 1977; Swales, 1979) and there has been more recently an endorsement at central level of the need for such comprehensive reports (DES, 1984).

One of the more controversial features of these newer reports is that they include structured reporting on attitudes and personality characteristics. There is much debate as to whether these kinds of assessments should be included in the final report, whether issued by the individual school, by the local authority, or by an examining body; and to what extent they should be included in the assessments and reports made during the course of schooling.

In this paper I am addressing this issue - reporting attitudes and personality characteristics. I am drawing from three Scottish studies. Two of them were studies of teachers' positions on these issues and one of parents' expectations. The first study is a national survey of teacher attitudes; the second is a study of teachers involved in a developmental project; and the third is a study of parental perceptions of such reports. The first teacher study I will discuss and the parents' study were funded by the then Social Science Research Council. The second teacher study was funded by the Scottish Education Department.

The teacher studies

The teacher survey was a study of teachers' response to two major documents

in Scotland, the Munn Report (SED, 1977a) and the Dunning Report (SED, 1977b). These reports are the basis of extensive reforms of secondary education in Scotland which are still in the course of implementation. Shortly after the reports were issued, the Scottish Council for Research in Education carried out a study of teachers' responses to the range of recommendations in the reports (Forsyth & Dockrell, 1979). Questionnaires were sent to a one-in-three sample of Scottish secondary schools. There was a response rate of just over 60%.

The Munn Report, which was concerned with curriculum, argued that among the aims of the schools were those 'concerned with the affective development of pupils. In educating young people it seems irresponsible to ignore their emotional and moral natures, or to assume that the educational process should not concern itself with their attitudes and values and whatever it is within human personality that predisposes people to act in particular ways' (SED, 1977a, p22). The Dunning Report, on assessment, recommended that a standardised, comprehensive record be kept of pupil performance, including attitudes.

Curriculum

Table 4.1: Curriculum and Use of Assessment in the Affective Domain

	headteachers		*other teachers*	
	% *yes*	% *no*	% *yes*	% *no*
Aims include affective development:	92	8	84	16
Curriculum to include:				
1. RELIGION	72	28	55	45
2. MORALITY	78	22	65	35
3. COMMUNITY SERVICE	48	52	50	50
Assessment to include affective characteristics:				
1. for GUIDANCE	93	7	88	12
2. for SCHOOL CERTIFICATE	42	58	44	56
3. for NATIONAL CERTIFICATE	44	56	45	55
4. for REFERENCES	90	10	80	20

How did the teachers respond to these various recommendations (Table 4.1)? The headteachers were virtually unanimous in their support of the Munn assertion that the aims of education must include the affective development of pupils: 92% of them supported the statement and fewer than 4% opposed it. Teachers were a little less certain: 84% of them endorsed this aim, 7% of them opposed it and, as with the headteachers, there was a small percentage who did not know.

One of the Committee's recommendations was that all pupils would follow eight modes of study which included religious studies and morality. There was less certainty about this recommendation: 72% of the headteachers endorsed the teaching of religion but only 55% of classroom teachers; slightly more than 78% of heads and 65% of teachers endorsed morality as a mode of study for all pupils.

One suggestion relating to affective development was that all pupils be required to take part in community service. Only 48% of heads and a bare 50% of classroom teachers were in favour of this recommendation.

Assessment and reporting

When it came to assessment and reporting, the focus of the study was on a standardised, comprehensive record including assessments of affective characteristics. 93% of heads and 88% of classroom teachers endorsed the compilation of such a record and its use by the school for curricular and vocational guidance. So the assessments were to be made.

When it came to issuing school leaving reports, however, there was a sharp division of opinions. Two options were offered: one was a certificate issued by the school and the other was inclusion in the national certificate as an endorsement made by the schools. Fewer than half of the heads supported either of these recommendations, 42% agreeing that the school itself should issue a certificate and 44% advocating endorsement of the national certificate. About the same proportion of classroom teachers was in favour of both options, 44% and 45% respectively, but, in all cases, fewer than half of the school staff thought such a use of their assessments was appropriate.

The final question in this section of the questionnaire referred to the use by the schools of these assessments for writing character references. This option was heavily endorsed, 90% of head teachers and 80% of classroom teachers approving the use of affective assessments for this purpose.

The vast majority of teachers accepted their responsibility for the affective development of their pupils and for assessing affective characteristics but fewer than half accepted the desirability of including such assessments on a leaving certificate.

The views of teachers involved in development work

That study was concerned only with school leaving certificates. Unfortunately we do not have teachers' reactions to the use of these assessments in reports during the course of schooling. An earlier study, however, had shown that teachers who were involved in the development of such assessments were overwhelmingly in favour of their use for reports during the course of schooling (Dockrell & Broadfoot, 1977). As Table 4.2 shows, the assessment and reporting of perseverance, interest, reliability, effort and carefulness were endorsed by over 75% of these teachers; and over 50% endorsed the reporting of other characteristics, including initiative, acceptance of discipline, willingness to help others, responsibility, confidence and self-reliance. Let me emphasise that this was a group that had been involved in development, and not a random sample of all teachers. We cannot be sure that the position of this group would be shared by others. What we can say is that when teachers are involved in these kinds of assessments they see their value for reporting.

Table 4.2: Teachers' Views of the Desirability of Including Characteristics in Reports

Characteristic	*% of all teachers in favour of inclusion in profile*
Interest	83
Perseverance	85
Reliability	77
Effort	77
Acceptance of discipline	74
Carefulness	76
Enterprise/Initiative	72
Willingness to help other people	64
Responsibility	60

This group, too, thought that assessments of affective characteristics should be included in the leaving report. 71% of those studied endorsed their inclusion (Dockrell & Broadfoot, 1977, p83). There was, however, an interesting division of opinion between the classroom teachers and the heads on the form of which the report should take: 90% of the heads favoured a letter or number grade but 68% of classroom teachers preferred comments only for this kind of assessment.

The study of parents

The third study I want to draw on was of what parents, pupils and employers wanted from teachers' assessments and school reports (McKay & Dockrell, 1983). Here I am reporting only parents' views of assessment in the affective domain.

This investigation was an intensive study of four secondary schools, each of them operating a different approach to reporting. We interviewed a stratified sample of parents. We stratified by pupil achievement and drew randomly within each stratum. In this way we were able to get the views of parents whose children were doing badly at school as well as those of parents of children who were doing well. It is not possible to make generalisations from ethnographic research of this kind. We cannot assert that 'n'% of all parents want this or that. We only know what the parents we interviewed wanted. This study does, however, provide an understanding of what parents want and why they want it. The interview schedule was made up of two sets of questions, a common set put to parents in all four schools and a second set relating to the specific practices in each school. Some attempt was made in the construction of the schedules to disentangle assessment and reporting but to do so entirely would be artificial. Reporting, after all, is an end product of assessment.

The first general point I would like to make is that the wide-spread assumption, particularly among teachers, that many parents are apathetic about their children's schooling was not borne out by our experience. The perceived lack of parental response may instead be a consequence not of apathy, but of schools telling parents the wrong things in the wrong way. It would hardly be surprising if schools were to get it wrong, for there is little evidence about parents' expectations of school reports. Schools continue to do what they have always done in reporting or attempting change with little systematic knowledge of how parents interpret reports, what they would like to receive in them and how they would expect to use them. All the parents in our sample were keen to make whatever contribution they could to their children's schooling, irrespective of the youngsters' current academic performance. Parents were, however, at a loss to know how they could make such a contribution. They looked to the school report to provide them with some kind of basis for action and for expert or professional advice, but felt that they did not usually receive it.

In response to the bald question: What kind of information would you like to receive in school reports? parents made it clear that they wanted information in two main areas, broadly cognitive (relating to subject performance) and non-cognitive (any other factors influencing subject performance or general development). This general finding is not surprising. What is significant is the nature of the specific items suggested within the areas.

There were three kinds of non-cognitive information that parents wanted:

first, information about attitudes such as effort, enterprise, interest, co-operation and so on, which are related to attainment; second, information about some aspects of personality, for example shyness or friendliness; and, finally, information about behaviour, in effect, conformity to school regulations.

Behaviour

All parents wanted the last kind of information. They expected the school to contact them immediately if there were any serious problems. Most parents did, however, accept that there would be minor deviations which the school could, and should, deal with adequately itself.

Attitudes

The majority of parents favoured the assessment of attitudes. They did so for several reasons. They thought that teachers' assessments would help them to get to know their pupils better and that such assessments would facilitate classroom management, enabling swift corrective disciplinary measures to be taken. They also thought that the development of healthy attitudes towards other people and to work was part of the teachers' job. They believed too that the assessment of attitudes would be helpful to them as parents. It would improve parents' knowledge of their own children by providing a different perspective. Parents who were in favour of the assessment of attitude by teachers seemed to assume that the assessment and development of pupils' attitudes could not be divorced from the process of teaching, and that social education was a joint responsibility of the home and the school.

School assessments had value because teachers had a wide experience of children and therefore had a broader basis for judgements than parents. Teachers also had a professional competence in making this kind of judgement. It is significant that all of the arguments in favour of the assessment of attitude are formative in nature, that is, to provide information to parents and teachers so that they may guide the development of pupils. Nowhere did we find summative assessment, for example reporting for selection or for references to employers, offered as a justification for the assessment of attitudes. The minority who were opposed to the assessment of attitudes doubted teachers' competence, and were aware of the limited opportunities for observation provided in the classroom.

When it came to reporting, parents wanted information which referred to their particular child, that is, they were not satisfied with a limited range of characteristics with a letter grade. They were looking for something which was individual. About half wanted a written comment in a report. The other half suggested that attitude assessments might better be reported to them orally at a parents' evening or during an interview with the guidance staff.

Personality

Questions about the assessment of personality were put only to the parents in a school using the SCRE profile assessment system and one using their own derivative of it. Even in these cases, where the parents were receiving such reports, it was necessary sometimes to prompt the parents by giving examples of the personality traits which were currently being assessed. A majority of the parents was in favour of the idea but there was the same polarisation of views as with attitude. The proportion in favour was smaller than with respect to attitudes.

The reasons for wanting these assessments were the same as with attitudes, with one interesting addition which was that a report on a pupil's character or personality might help the pupil to get a job, presumably on the assumption that such a report would be favourable. This is interesting in that it is the first justification offered which might be termed summative in nature. Those who were opposed to such assessments held that these aspects of personality were not prompted or developed by the school and indeed could not be.

As with attitude, parents were against the idea that assessments of personality should be norm-referenced, arguing that teachers should match the qualities observed against certain standards which they themselves held. There was a feeling that letter grades or ticked descriptions of non-cognitive characteristics are insufficient and that these should be replaced or supplemented by written comment. Such comment is more personal and therefore more appropriate for conveying information of a personal nature and in addition makes parents feel that the report is about their individual child. They believed that when teachers were faced with a blank space to be filled with a written comment they had to think about the individual. Parents also believed that grades, symbols and ticked boxes were not sufficiently flexible to cope with the subtleties of personality.

If reports were to contain only one assessment of any non-cognitive characteristic, then parents would prefer that to be based on the consensus view of all, or at least some, of the pupils' teachers. Some parents stated a preference for receiving individual assessment from each teacher, arguing that this could reveal interesting patterns and exceptions to patterns. Parents, in general, were concerned that school records of non-cognitive characteristics should be updated often, especially when improvement was shown.

It is clear that when parents are offered a more comprehensive reporting system than is currently the practice in Scotland, most of them are pleased to get it. Most parents, whether their children are attending denominational or non-denominational schools, think of the schools as partners in the total education of their children and not merely as institutions for imparting knowledge and skills.

Conclusions

There are, I think, a number of significant conclusions to be drawn from these sets of findings. The first is that teachers accept that schools have responsibility for the affective development of pupils and that many parents, but not all, share this position. Education, in its affective aspects as well as in its cognitive aspects, is a joint enterprise by home and school. When it comes to reporting, there is understandable trepidation on the part of teachers, and uncertainty on the part of parents. What might be desirable in principle might be difficult, and indeed questionable, in practice. However, both teachers and parents who have had experience of a carefully constructed system of assessment and reporting in this crucial but difficult area are generally favourable. That does not mean to say, of course, that all systems will be acceptable to all teachers and parents, far from it. There seems to be a substantial minority, both of parents and teachers, who have reservations about, or are indeed opposed to, such assessments at all.

Comment

Let me end with some personal comment.

Schools have some responsibility for the social development of their pupils. Their aims should be explicit. They should be clear to students and parents, and they should be part of the formative assessment process.

As for summative purposes, schools have little choice. The forms that are sent out to schools now by some employers require these assessments. Are teachers going to refuse to fill in these forms? Are parents going to say that teachers should not fill them in? The consequences for youngsters where the forms are not completed might well be very serious indeed. If these assessments are going to be made, they must be well made. We need a carefully constructed system which will ensure that assessments are comparable.

Finally, I would argue that if these assessments are to be reported to employers, or anyone else for that matter, they should be known to the young people and their parents beforehand, though not necessarily included on a leaving certificate.

5

Certifying School Graduates

This paper first appeared in 1981 in 'Evaluation Roles in Education', a collection of papers on the topic of evaluation, edited by Arieh Lewy and David Nevo of Tel-Aviv University, published by Gordon and Breach.

School leaving certificates are a nearly universal phenomemon. Whether it is the High School Diploma of the United States, the *Abitur* of Germany, the *Slutbetyg* of Scandinavia or the British School Certificate, the practice is widespread and even where there are criticisms of the existing form of certificate, as, for example, in Australia, a certificate of some kind has typically been retained. However diverse the procedures involved, there is a general recognition of the transition from the third to the fourth of Shakespeare's 'seven ages of man', the successful completion of one stage and progression to the next. In many countries the gaining of the certificate is associated with ritual and celebration that suggests that it has been elevated to the status of a 'rite of passage'.

These rituals take an elaborate form in the United States. The graduation ceremony and the graduation ball are the culmination of high school years. In other countries the celebration may be more modest as in the British prizegiving, or even more elaborate, if the cinema is to be believed, in Sweden. It is the ceremonial surrounding the event which suggests that this is not simply a necessary routine, like taking a college entrance examination or sitting for an Open Scholarship, but something of greater social and psychological significance, like a Bar Mitzvah or a military Passing Out parade.

Graduation is an important event in the life of youngsters in many societies, not for what it records, but for what it presages, a new status in the adult community. It is important to recognise this ritual significance of the leaving certificate for it is easy to point out the limitations of certificates for other purposes. They are unsatisfactory for many, if not all, of their ostensible purposes, yet the criticisms do not seem to come to the heart of the concerns of the consumers, pupils, parents and teachers. None of the criticisms takes account of the emotional significance of graduation.

Uses of school certificate

Formally, a school leaving certificate is merely a record of past achievement. If this were all it might have interest for the pupil, but little significance. Ostensibly the certificate simply indicates that the individual has completed a

defined stage and usually that a satisfactory standard has been reached. However, it is frequently seen by pupils as indicating something permanent and absolute, like passing a driving test, more a measure of height than of weight. Sometimes adults produce school leaving certificates obtained many years previously as indicating a level of competence; even though they may accept that the curriculum has changed or that they may need to brush up some aspects of study.

The importance of the certificate for pupils lies in the fact that it is seen by them as a guide for future action. At its simplest, a satisfactory grade may be taken as demonstrating a sufficient level of competence so that the pupil need not concern himself further with the study of that particular subject, or it may be taken to indicate that he is now sufficiently competent in some subject to move on to a higher and more difficult level of study. Certificate marks may be a source of more specific guidance too. If the certificate contains a higher grade than expected in one area and a lower grade than expected in others, it may be taken to suggest a change of programme. It is seen as an indication that the youngster is 'not good at' some subject or group of subjects. This conclusion may be drawn in spite of the cumulative evidence of years of experience with the subjects in the ordinary school setting. The pupil's perception of the certificate is as important as the formal constraints which may be evoked in a particular society by employers or tertiary education institutions. Parents' perceptions are usually similar and similarly confused. In one recent study (Ryrie, Furst and Lauder, 1979), parents saw the certificate as a judgement of their children. Some saw it as a judgement of their children's ability. 'He was never good at school' was a phrase parents used as an explanation, almost an extenuation, of performance in the leaving certificate. Others seemed to see the certificate as a judgement of their children's efforts, either in general or in specific areas. 'He did not work at his maths, he never liked it', was an explanation offered for failure in a particular subject.

The certificate was also seen by parents as a guide to action. The most frequent advice given by parents to pupils was 'to do what you are best at'. Educationally more sophisticated parents are able to recognise the limits of the certificate and to advise their children about the effects on their level of attainment of particular programmes or courses that they have taken and of the effects of the school they have attended and the teachers who have taught them. Lack of success recorded in the certificate is not seen by these parents as having any permanent significance, simply as indicating achievement at a given point in time and in specific circumstances. Nonetheless the majority of both pupils and parents see the certificate not simply as a record of achievement but as a statement about ability and ultimate levels of achievement.

Employers frequently take the leaving certificate to indicate both a general level of competence and a mastery of a specific body of knowledge or set of

skills. Advertisements for jobs for school leavers frequently specify they are looking for high school graduates or holders of the leaving certificate. For certain jobs, passes in designated subjects like mathematics or science or a foreign language are required, even though this knowledge may not be relevant to the employer's needs. Here the employer is using the certificate as a sieve to select those who have demonstrated a general level of competence and the specification of passes in certain subjects is not related to the needs of the job concerned but to the expectation that they indicate a higher level of general competence.

Sometimes examination results may be seen in a negative way. Some employers say that they are looking for a demonstrated lack of academic success among potential employees on the assumption that they will be more satisfactory for undemanding and routine tasks. However, no advertisements have been noted that state that applicants should not possess a leaving certificate.

Some employers, either individually or as industrial groups, set their own examinations or use those designed by psychological test bureau. They regard possession of a leaving certificate, or lack of it, as irrelevant to their particular employment. They prefer tests which give evidence of mechanical or clerical skills which would serve as a basis for specific training. Nevertheless, a substantial number of employers specify the possession of a school leaving certificate as a minimum basis for consideration for employment.

Teachers use the certificate, both as a stick and a carrot. Threats of failure are frequently used to goad the less successful pupils, and the promise of success to encourage the more successful to even greater efforts. Some teachers of course dismiss this kind of motivation as artificial and say that the intrinsic interest of the subject itself or its obvious relevance for some later programme should be sufficient motivation. It is, however, a widespread belief among teachers that a formal target provides a stimulus and a motivation. At the least it encourages students to stay at school to the completion of the course and the award of the certificate. There is great variety among the attitudes of tertiary education institutions to school leaving certificates. In the most extreme form some institutions make no reference to a diploma or certificate at all. Open admission typically applies to vocational courses of low academic demand or to mature students of a defined age or who have completed a specified number of years in industry. At the other extreme are the institutions which require a high school diploma or certificate only and guarantee admission to all those who hold this qualification. This is the practice in many parts of the United States and, until recently at least, in the United Kingdom. The standard of the diploma is different in the two societies. In the United States as many as 80% of the age group may have the high school diploma which qualifies to proceed to the next level, whereas in England, it is only 20% or so of the population who obtain the

certificate at the required standard. In both cases, however, the obtaining of the certificate is the guarantee of admission to higher education.

In some countries where the school leaving certificate is not taken into account for admission to tertiary education it is because the institutions, either individually or collectively, have an entrance examination of their own as in Japan. More common is the combination of school assessments and a college entrance examination as in many parts of the United States. In Eastern Europe, too, many tertiary institutions require both a school leaving certificate and a satisfactory standard in an entrance examination.

The form of the certificate

As striking as the diversity in expectation is the diversity in the basis for the award of the certificate. The certificates all include assessments of performance in school subjects. The pattern seems to have been set nearly four hundred years ago in the ordinances of the various German States, codified by Frederick the Great in 1776 and by Napoleon in 1808 (Hotyat, 1962). Whether or not the certificate strays outside the strictly academic boundary varies greatly. Though the list is usually restricted to conventional subjects it sometimes includes, as in Scandinavia, aesthetic subjects, technical subjects and physical education.

Some certificates record each of the subjects in which a student obtained a satisfactory level. In other countries, in order to obtain a certificate at all a group of subjects must be passed. These subjects cover all or a selection of the school curriculum. They may require a pupil, for example, to pass in the mother tongue, mathematics and any four or five other subjects or they may be more specific and require passes in one or more subjects from each of several designated curriculum areas. They may require a pass in a science subject but not specify chemistry or physics and in a social studies subject but not specify whether it be history or geography. In Britain there has been a move from the group certificate which required passes in specified subjects or groups of subjects to a simple record of the subject or subjects where a satisfactory level was achieved.

What is recognised as a satisfactory performance varies from country to country. In its simplest form the certificate simply records a pass/fail indicating that the student has reached a satisfactory standard in the subjects listed. Others are more elaborate and give a percentage or grade according to some established system. Some refer to a single level of attainment while others recognise performance at more than one level: Higher and Ordinary levels in Scotland. An advanced level of performance may be required in two or three subjects and a lower level in two or more others. It is not only the countries which require a group of subjects which recognise different levels. It is possible in Scotland, for example, to obtain a certificate that records a pass in a single subject at the Higher level or any combination of Higher and Ordinary grade.

In countries where a certificate records passes at different levels the level of pass may be important. A pass at Ordinary level in a subject may be acceptable for admission to tertiary education depending on its role in the student's programme. Thus, for admission to university, all students may be required to have a pass in the mother tongue at the Ordinary level, but would be required to have a pass at Higher level if they wished to study the mother tongue at university.

In some countries it is the average mark that counts. In the United States marks are added together and averaged to give a 'grade point average' and thus a position in the total graduating group from a particular high school. In Sweden, marks are added together to give a grade point average which is taken to indicate a position in a national graduating group. A somewhat similar process is followed for university admission in the United Kingdom. There, marks at particular levels are given points and added together to give a total number of points.

This procedure has certain assumptions which are not easily met. One of them is that the marks are of the same importance or can be equated. They might need to be weighted either by the duration of the course or of the importance of the subject in the programme. In Sweden, marks from one year, two year, three year and four year courses are added together, unweighted, to give a total number of points. Consequently, a student following a predominantly scientific programme may find that his four year course in mathematics or physics contributes no more to his final average than a one year course in an aesthetic subject. Another assumption is that different subjects at the same level are of equal difficulty. In the United Kingdom the marks are from an external examination and are taken to be comparable. In Sweden, comparability is obtained by assigning to schools a distribution of marks based on the results of an external monitoring examination. In other countries there is no means of comparing marks.

There has, in recent years, been some discontent with the exclusive focus on cognitive achievement, even when this is broadly defined to include aesthetic and technical subjects and physical education. In the United Kingdom alternative systems have been developed and have been considered widely, if not generally adopted. One of these is the record of personal achievements developed in Swindon in England (Swindon Education Committee).

This report allows the student to have included in his record all those aspects of his total performance which he believes to be significant.

A somewhat more structured approach was followed in Scotland in the development of the Pupil Profile Assessment System (SCRE, 1977). In this system the final report includes not only the traditional statement of achievement in all aspects of the curriculum, but also an assessment of general

skills, i.e. listening, speaking, reading, writing, visual understanding and expression, the use of number, physical co-ordination and manual dexterity.

An example of a completed school leaving report is presented in Figure 5.1.

Each of these assessments is made on a four point scale where reference is made to a defined standard, so that the report card carries a descriptive phrase for each of four standards in the eight areas. In listening, for example, the four standards are:

(1) acts independently and intelligently on complex verbal instructions;
(2) can interpret and act on most complex instructions;
(3) can interpret and act on straightforward instructions;
(4) can carry out simple instructions with supervision.

In physical co-ordination the four standards are:

(1) has natural flair for complex tasks;
(2) has mastery of a wide variety of movements;
(3) can perform satisfactorily most everyday movements;
(4) can perform single physical skills such as lifting or climbing.

The teachers are given manuals appropriate to their particular subjects indicating what kind of behaviours would merit a mark at each specified standard. The assessments of general skills are gathered from all teachers and pooled. Teachers only report on those skills that they have an opportunity to observe. Most teachers, for example, can assess a pupil in listening and speaking, but it is teachers of geography and art who are most likely to be able to make assessments in visual understanding and expression.

This system also includes assessments of two affective characteristics - enterprise and perseverance. Here, too, a series of guides have been developed for teachers. Behavioural examples, called crucial indices in this system, have been developed for each school subject. In the case of English teachers the indices at each level are as follow:

Conscientiousness/Perseverance
- Completes work only if teacher stands over him/her
- Often forgets to do homework
- Continually asks questions about what to do
- Interested in most work but is not prepared to work independently
- Carefully corrects mistakes
- Attempts difficult work and does not give up easily

Confidence
- Afraid to write anything down in case it is wrong

- Never speaks in class except when the atmosphere is extremely informal
- Prefers to work quietly rather than ask questions of the teacher
- Answers simple questions but prefers not to try more difficult ones
- Gives answers based on own experience
- Speaks out his/her own opinions

A greater emphasis is given to affective characteristics in a leaving report used by schools in the Lothian Region of Scotland. Like the 'pupil profile' system this report includes a record of academic achievement, of some basic skills and of three affective characteristics, attitudes to school work, relations to teachers and relations with other pupils.

The inclusion of affective characteristics on the Certificate is a controversial issue. A recent survey of secondary teachers in Scotland (Forsyth & Dockrell, 1979) showed that 90% believe that these assessments should be made and noted, but rather less than half were in favour of including them in a school leaving certificate. Approximately the same proportion thought that the assessment should not be formally recorded on a certificate but be used exclusively for the preparation of references and the completion of forms.

In practice such assessments are asked of many schools by prospective employers and in the United Kingdom for admission to university on the standard admission form. A certificate of this kind was standard practice in Norway until recently for transfer from lower to upper secondary school.

In spite of the almost universal prevalence of a certificate there is considerable variation in the basis for the certificates. It is difficult to relate these variations to either educational or economic circumstances. Countries with apparently comparable situations have widely different practices. It is not obvious why Germany should have a predominantly school based assessment system and France have a largely external one.

In France, the *Baccalauréat* is an external examination as is the General Certificate of Education in England. In the United States and Japan, each school awards its own diploma and the awarding of grades is carried out on a purely internal basis. Intermediate between these two extremes is the situation in Holland, where school based assessments and external examination marks are combined.

Internal and external assessment

Sharply contrasted as the systems may seem, all have some element of both internal and external assessment. In Germany, the Abitur is based on the marks awarded by the pupils' own teachers, but these marks are usually monitored by a colleague and may in some circumstances involve external moderation. The Scottish Certificate of Education is formally an external examination, set and marked by an Examination Board, but schools are required to prepare and

OTHER OBSERVATIONS

(includes other school activities, other awards and comments on positive personal qualities)

Royal Life-Saving Society - Bronze Medallion

Member School Photographic Club, School Debating Society

Member of School Skiing trip to Austria Jan 1976

She has been resourceful in finding costumes for the school play.

She has recently shown an appreciation and enjoyment of literature and has read widely outside the syllabus.

Works well on group activities; gets on well with both pupils and teachers. Readily accepts responsibility, particularly in social activities.

Notes

The grades A - D represent approximately 25% of the year group in each case.

The skill gradings represent a consensus derived from the individual ratings of each teacher's knowledge and reflect the standard obtained by the pupil with reasonable consistency.

All the information contained in this report is based on profile assessments contributed by each teacher on a continuous and cumulative basis, including observations of personal qualities and informal activities.

© 1976 SCRE

SCHOOL LEAVING REPORT

This is a brief report on Queenie Quarry

Date of Birth 13/7/1960

who completed class S4

in Tanochbrae High School

and left on 3rd July 1976

This report is the result of continuous assessment by all the teachers of this pupil and has the authority of

E P Smith Head Teacher

J McGregor Director

(pages 1 and 4)

Table 5.1: School Leaving Report – Pupil Profile Assessment System

SKILLS

LISTENING

- Acts independently and intelligently on complex verbal instructions ☐
- Can interpret and act on most complex instructions TH
- Can interpret and act on straightforward instructions ☐
- Can carry out simple instructions with supervision ☐

SPEAKING

- Can debate a point of view ☐
- Can make a clear and accurate oral report TH
- Can describe events orally ☐
- Can communicate adequately at conversation level ☐

READING

- Understands all appropriate written material TH
- Understands the content and implications of most writing if simply expressed ☐
- Understands uncomplicated ideas expressed in simple language ☐
- Can read most everyday information such as notices or simple instructions ☐

WRITING

- Can argue a point of view in writing ☐
- Can write a clear and accurate report TH
- Can write a simple account or letter ☐
- Can write simple messages and instructions ☐

VISUAL UNDERSTANDING AND EXPRESSION

- Can communicate complex visual concepts readily and appropriately ☐
- Can give a clear explanation by sketches and diagrams ☐
- Can interpret a variety of visual displays such as graphs or train timetables TH
- Can interpret single visual displays such as road signs or outline maps ☐

USE OF NUMBER

- Quick and accurate in complicated or unfamiliar calculations TH
- Can do familiar or straightforward calculations, more slowly if complex ☐
- Can handle routine calculations with practice ☐
- Can do simple whole number calculations such as giving change ☐

PHYSICAL CO-ORDINATION

- A natural flair for complex tasks ☐
- Mastery of a wide variety of movements ☐
- Can perform satisfactorily most everyday movements TH
- Can perform simple physical skills such as lifting or climbing ☐

MANUAL DEXTERITY

- Has fine control of complex tools and equipment ☐
- Satisfactory use of most tools and equipment TH
- Can achieve simple tasks such as wiring a plug ☐
- Can use simple tools, instruments and machines such as a screwdriver ☐

SUBJECT/ACTIVITY ASSESSMENT

Curriculum Area	Subjects Studied (includes final year level where relevant)	Years of Study	Achievement	Enterprise (includes flair, creativity)	Perseverance (includes reliability, carefulness)
Aesthetic Subjects	Drawing	1-4	2	2	1
	Music	1-4	2	3	3
Business Studies					
Community/ Leisure Activities	Social Education	1-4	3	2	3
Crafts	Pottery	3-4	2	1	3
English	English	1-4	2	1	3
Mathematics	Arithmetic	1-4	1	1	2
Other Languages	German	2-4	2	2	3
Outdoor Studies	Outdoor Pursuits	3-4	2	2	3
Physical Education	General	1-4	3	1	3
Science	Biology	3-4	1	2	2
Social Subjects	History	1-4	2	1	3

(pages 2 and 3)

submit to the Board an order of merit. This order of merit can be used as grounds for appeal, if a student's performance is markedly lower than anticipated. Similarly, in France, school marks may be used in cases of illness or in borderline cases. In the United States there is a system of external examination, parallel to the high school diploma. The Examination Boards set achievement tests which are called Advanced Placement Examinations. As the name implies, they are intended to be at a standard in advance of that of the normal marks recorded in the high school diploma. New York is an exception to the usual American pattern, where there is a Board of Regents comparable to the external examining boards in the United Kingdom.

In the United Kingdom the Certificate of Secondary Education has three modes. These modes range from one which involves simply conventional external examination to a system which relies exclusively on school based assessments which are reviewed by external moderators.

When the school leaving age was raised in Britain so that all pupils were required to stay at school to the point at which an external school leaving certificate was issued, a number of problems arose. To provide for the needs of pupils whose level of attainment was lower than that traditionally assessed by the General Certificate of Education an additional Certificate of Secondary Education was established. These certificates were not mutually exclusive and indeed were designed to have an overlap so that a pass at a satisfactory level in either examination was acceptable for progress to the next stage of education. The General Certificate of Education Examination Boards, however, had a long tradition behind them and had well established practices for the preparation of curricula and for assessment. The new Boards had no such traditions and were free to experiment. Indeed, they had to experiment, because there were no curriculum guidelines for the pupils they would be examining and the schools were encouraged to develop their own.

The new Boards developed three modes of assessment. Mode 1 was the traditional procedure where the Examination Board prepared a curriculum and set and marked externally an examination. Mode 2 was where a school, or more usually a consortium of schools, developed a curriculum which was then examined in the traditional way, that is, by external examination set and marked by the Examining Board. Mode 3 was the most unusual from the assessment point of view. The schools were not only responsible either individually or more usually in consortia for the development of curricula but were responsible for their assessments. Given the British tradition of the external examination, some procedure was necessary to ensure that the internal assessments made in Mode 3 were equivalent to those made under the more traditional approach and indeed that the assessments made by different schools were comparable. For this purpose a system of moderation was developed.

The Mode 3 procedure was time-consuming. First, the schools prepared

curricula which were submitted to the Certificate Boards. After initial review by the Board there were discussions between the teachers and representatives of the Board, who might be permanent employees of the Board or senior teachers from other schools, to discuss the suitability and acceptability of the proposals. At this stage the curricula were frequently revised, sometimes substantially.

As part of their submission the schools had to say how achievement would be assessed. It could be either in the form of traditional examinations which would be marked internally or it could be on the basis of exercises, practical or formal or any combination of these. Only when the moderators were satisfied did the Board accept the curriculum for certification.

Records of pupils' work had to be retained so that they could be re-marked by a panel of moderators. For small groups of students, usually less than 20, the work of all pupils was re-assessed by the external examiner. For larger groups a sample was usually considered sufficient. The moderator typically had three concerns. First, that the order of the pupils was correct. Second, that the spread of grades awarded was appropriate, and third, that grades awarded corresponded generally to those awarded by other schools or by other means.

There has been much confusion, particularly among parents and employers, about the meaning of the new certificate. Much controversy and indeed some rejection by tertiary education institutions of assessment made exclusively by teachers. Nonetheless, Mode 3 has been generally welcomed by teachers though only a minority choose to prepare pupils for this kind of assessment.

In Sweden, internal and external marks are combined in an unusual way. The range of marks that a teacher should assign to his class is determined by an external examination set not at the end of the school program but sometime during its course. The purpose of this examination is not to decide the marks of the individual pupil but to prescribe the range of marks that may be awarded by the school as a whole.

Final examinations versus cumulative records

Where a school component is taken into account the basis for it varies. In some cases reliance is placed primarily on a final examination, set and marked by the student's own teacher. There is a close parallel here to the external examination set by the Examination Board. In other cases a cumulative record is kept of performance during the course and this is taken into account as in Germany, when the marks obtained in the *Klassenarbeiten* are combined with examination results. In Britain, as noted above, the Certificate of Secondary Education assessment may consist entirely of assessments of pupils' work during the period of instruction or may include final examination marks as well.

This system of cumulative assessment became very popular in the United

Kingdom some ten years ago. Students particularly regarded it as fairer than a single assessment at the end of the course. This view has been subject to some revision recently as it was recognised that a single and untypical poor performance during the course might result in a lower average mark than seemed justified. While final examinations have been criticised for the pressure they put on students some have complained of the continuing pressure from a system of cumulative assessments.

Certificates and the examinations on which they are based have been a subject of research for many years. In 1888 Professor F Y Edgeworth published an article on the statistics of examinations (Hartog, 1918). In that and other papers Edgeworth not only defined the true mark, as we now use the term, but also outlined the major sources of error and their likely contribution to the total error inherent in a typical examination. The most extensive and systematic early studies of school leaving certificates were those of Hartog and Rhodes in the thirties in particular their study of the marks awarded to the same papers by different examiners (Hartog, Rhodes and Burt (1936). These studies were part of an international series of studies conducted in England, Finland, France, Germany, Norway, Scotland, Sweden and the United States under auspices of the Carnegie Corporation.

These early findings have been replicated many times since and summarised by Ingenkamp (1977). In German speaking countries, where the assessments are largely internal, the correlations between marks in the *Abitur* and success in universities range from 0.06 to 0.49. Similar results have been found in the United Kingdom where external examinations were the basis of prediction. Entwistle, Nisbet, Entwistle & Cowell - (1971) reported a correlation of 0.32 between the results of GCE and academic success and Powell (1973) in a comprehensive Scottish study reported correlations from 0.18 in the Faculty of Arts, to 0.41 in the Faculty of Engineering.

Correlations between vocational success and leaving certificate results are no higher. In a series of studies conducted in Scotland (Ryrie & Weir, 1978) a number of significant correlations emerged between School Leaving Certificates and success in vocationally oriented programmes but none of them exceeded 0.29. These findings were in harmony with those of other researchers, perhaps because 'the apprenticeship process would seem over 4 years in the lives of young adults to produce such variations in performance as to throw doubts on the purposes of attempting to predict success' (*ibid*, p158).

There have, over the years, been investigations into school leaving certificates by national committees, comparative studies of practices in various countries (Hotyat, 1962; McGuire, 1976) and analyses of the consequences of different approaches (Elley & Livingstone, 1972) and reforms in some countries, but the issuing of School Leaving Certificates remains a universal or near universal phenomenon and the practices of each country seem remarkably impervious to change.

REFERENCES

ASSESSMENT OF PERFORMANCE UNIT (1977) *Why, What and How.* Department of Education and Science.

AUSUBEL, D. P. (1967) Learning theory and classroom practice, *Ontario Institute for Studies in Education Bulletin*, No.1.

AUSUBEL, D. P. *and* FITZGERALD, D. (1962) Organizer, general background and antecedent learning variables in sequential verbal learning, *Journal of Educational Psychology*, 53, 243-9.

BEREITER, C. *and* ENGELMANN, S. (1966) *Teaching Disadvantaged Children in the Preschool.* Englewood Cliffs, N.J.: Prentice-Hall.

BLACK, H. D. (1983) *Constructing Diagnostic Instruments in Home Economics.* Edinburgh: The Scottish Council for Research in Education.

BLACK, H. D. *and* GORING, R. (1983) *A Diagnostic Resource in Geography.* Edinburgh: The Scottish Council for Research in Education.

BOARD OF EDUCATION (1926) *The Education of the Adolescent.* Report of the Consultative Committee to the Board of Education (The Haddow Report). London: HMSO.

BOARD OF EDUCATION (1938) *Secondary Education with Special Reference to Grammar Schools and Technical High Schools.* Report of the Consultative Committee on Secondary Education. (The Spens Report). London: HMSO.

BURT, *Sir* Cyril (1961) Intelligence and social mobility, *British Journal of Statistical Psychology*, 14, 3-24, May.

BURT, *Sir* Cyril (1966) The genetic determination of differences in intelligence. A study of monozygotic twins reared together and apart, *British Journal of Psychology*, 57, 137-53.

BURT, C. *and* HOWARD, M. (1957) The relative influence of heredity and environment on assessments of intelligence, *British Journal of Statistics Psychology*, 10, 33-63.

CAPLAN, N., MORRISON, A. *and* STANBAUGH, R. J. (1975) *The Use of Social Science Knowledge in Policy Decisions at the National Level.* Institute for Social Research, University of Michigan.

CARLSON, D. (1980) Statewide assessment in California, *Cadermos de Pesquisa*, December.

COOPER, R. M. *and* ZUBECK, J. P. (1958) Effects of enriched and restricted early environments on the learning ability of bright and dull rats, *Canadian Journal of Psychology*, 12, 159-64.

DAVIS, A. (1948) *Social Class Influence on Learning.* Cambridge, Mass.: Harvard University Press.

DE LANDSHEERE, G. (1979) *Dictionnaire de l'evaluation et de la recherche en education.* Paris: Presses Universitaires de France.

DEPARTMENT OF EDUCATION AND SCIENCE (1984) *Records of Achievement for School Leavers.* London: HMSO

DEVEREUX, A. (1979) What is education for - an industrialist's view. *CCC News 2*, February.

DOCKRELL, W. B. (1970) On Intelligence. In: Dockrell, W. B. (ed) *On Intelligence: contemporary theories and educational implications: a symposium, Toronto, 1969.* Toronto: The Ontario Institute for Studies in Education.

DOCKRELL, W. B. (1980) Teachers' assessments of their pupils' affective characteristics, *Enfance*, 1980, 4/5, pp295-299.

DOCKRELL, W. B. (1981) Certifying school graduates. In: Arieh Lewy *and* David Nevo (eds), *Evaluation Roles in Education.* London: Gordon and Breach.

DOCKRELL, W.B. (1982) The contribution of national surveys of achievement to policy formation. In: D. B. P. Kallen, G. B. Kosse, H. C. Wagenaar, J. J. J. Kloprogge *and* M. Vorbeck (eds), *Social Science Research and Public Policy-Making: A Reappraisal.* Windsor: NFER-Nelson.

DOCKRELL, W. B. (1985) Recent Developments in Assessment and Examinations. Paper given to the Annual Conference of the National Council on Measurement in Education.

DOCKRELL, W. B. *and* BROADFOOT, P., (1977) *Pupils in Profile*, Edinburgh: The Scottish Council for Research in Education.

DOCKRELL, W. B. *and* HAMILTON, D. (1980) *Rethinking Educational Research.* London: Hodder & Stoughton.

EAKER, R. *and* HUFFMAN, J. (1980) Interview, *Communication Quarterly*, 4, 1.

EGGLESTON, J. (1980) *School-Based Curriculum Development in Britain.* London: Routledge & Kegan Paul.

ELLEY, B. W. *and* LIVINGSTONE (1972) *External Examinations - Internal Assessments.* Wellington: New Zealand Council for Educational Research.

ENTWISTLE, N., NISBET, J., ENTWISTLE, P. and COWELL, M. (1971) The academic performance of students, *British Journal of Educational Psychology*, 41 (3), 258-278.

FORSYTH, J. P. *and* DOCKRELL, W. B. (1979) *Curriculum and Assessment: the response to Munn and Dunning.* Edinburgh: The Scottish Council for Research in Education.

GUILFORD, J. P. (1967) *The Nature of Human Intelligence.* New York: McGraw-Hill.

GUINAGH, J. P. (1969) An experimental study of basic learning ability and intelligence in low socio-economic populations. Unpublished doctoral dissertation. Michigan State University.

HAGGARD, E. A. (1954) *Social Status and Intelligence: An Experimental Study of Certain Cultural Determinants* (Genetic Psychological Monographs, Vol.XLIX, No. 2), Provincetown, Mass.: The Journal Press.

HALSEY, A. H. (1958) Genetics, social structure and intelligence, *British Journal of Sociology*, 9, 15-28.

HARTOG, P. J. (1918) *Examinations and their Relation to Culture and Efficiency.* London: Constable.

HARTOG, P. J., RHODES, E. C., *and* BURT, C. (1936) *The Marks of Examiners*. London: MacMillan.

HESS, R. D. *and* SHIPMAN, V. C. (1965) Early Experience and the socialization of cognitive modes in children. *Child Development*, 36, 869-86.

HOTYAT, F. (1962) *Les examens, les moyens, d'evaluation dans l'enseignment*. Paris: Institute of UNESCO.

HUNT, J. McV. (1961) *Intelligence and Experience*. New York: Ronald Press.

HUNT, J. McV. (1969) Has compensatory education failed? Has it been attempted? *Harvard Educational Review*, 39, 278-301, Spring, 1969.

HUSEN, T. (1967) *International Study of Achievement in Mathematics*. 2 volumes. Stockholm: Almqvist and Wiksell.

INGENKAMP, K. H. (1977) *Educational Assessment*. Slough: National Foundation for Educational Research.

JENSEN, A. R. (1969) How much can we boost I.Q. and scholastic achievement? *Harvard Educational Review*, 39, 1-123.

KAGAN, J. (1968) *On Cultural Deprivation in Environmental Influences*. Glass D. C. (ed). New York: Rockefeller University Press.

KAGAN, J. S. (1969) Inadequate evidence and illogical conclusion. *Harvard Educational Review*, 39, 274-8.

KALLEN, D. B. P., LOSSE, G. B., WAGENAAR, H. C., KLOPROGGE, J. J. J., *and* VORBECK, M. (eds) *Social Science Research and Public Policy-Making: a reapraisal*. Windsor: NFER-NELSON.

KOCH, S. (1969) Psychology cannot be a coherent science. *Psychology Today*. September.

LEWY, A. *and* NEVO, D. (eds) (1981) *Evaluation Roles in Education*. London: Gordon and Breach.

LURIA, A. R. (1961) *The Role of Speech in Regulation of Normal and Abnormal Behaviour*. London: Pergamon.

McGUIRE, J. (1976) *An Outline of Assessment Methods in Secondary Education in Selected Countries*. Edinburgh: The Scottish Council for Research in Education.

McKAY, N. *and* DOCKRELL, W. B. (1983) Parental Perceptions of School-based Assessment. Paper given at the Annual Conference of the British Educational Research Association, London.

MINISTRY OF EDUCATION (1965) *Children and their Primary Schools*. Report of the Central Advisory Council for Education (England). 2 volumes. (The Plowden Report). London: HMSO.

MULKAY, M. J. (1977) Sociology of the scientific research community. In: Spiegel-Rosing, I. *and* De Sola Price, D. (eds), *Science, Technology and Society*, Sage Publications.

OPPENHEIMER, R. (1955) Analogy in Science. *American Psychologist*, 10, 127-35.

PINKUS, J. (Ed) (1980) *Educational Evaluation in the Public Policy Setting*. Rand.

POWELL, J. L. (1973) *Selection for University in Scotland.* London: The University of London Press.

RYRIE, A. C., FURST, A. and LAUDER, M. (1979) *Choices and Chances.* Edinburgh: The Scottish Council for Research in Education.

RYRIE, A. C. *and* WEIR, A. D. (1978) *Getting a Trade.* Edinburgh: The Scottish Council for Research In Education.

SCOTTISH COUNCIL FOR RESEARCH IN EDUCATION (1963) *The Scottish Scholastic Survey, 1953.* London: University of London Press.

SCOTTISH COUNCIL FOR RESEARCH IN EDUCATION (1968) *Rising Standards in Scottish Primary Schools.* London: University of London Press.

SCOTTISH COUNCIL FOR RESEARCH IN EDUCATION (1969) *The Attainments of Scottish Ten Year Olds in English and Arithmetic.* London: University of London Press.

SCOTTISH COUNCIL FOR RESEARCH IN EDUCATION (1977) *Pupils in Profile: making the most of teachers' knowledge of pupils,* Edinburgh: The Scottish Council for Research in Education.

SCOTTISH COUNCIL FOR RESEARCH IN EDUCATION (1977) *SCRE Profile Assessment System Manual,* Edinburgh: The Scottish Council for Research in Education.

SCOTTISH EDUCATION DEPARTMENT (1977a) *The Structure of the Curriculum in the Third and Fourth Years of the Scottish Secondary School.* (The Munn Report). Edinburgh: HMSO.

SCOTTISH EDUCATION DEPARTMENT (1977b) *Assessment for All - a report of the Committee to Review Assessment in the Third and Fourth Years of Secondary Education in Scotland.* (The Dunning Report). Edinburgh: HMSO.

SCOTTISH EDUCATION DEPARTMENT (1977c) *The Educational System of Scotland.* Edinburgh: Scottish Information Office.

SCOTTISH EDUCATION DEPARTMENT (1980) *Learning and Teaching in Primary 4 and Primary 7.* A Report by HM Inspectors of Schools. Edinburgh: HMSO.

SIMON, B. (1960) *Studies in the History of Education, 1780-1920.* London: Lawrence and Wishart.

SKINNER, B F (1961) *Cumulative Record.* New York: Appleton-Century-Crofts, pp23-6 and 185-94.

SULLIVAN, E. V. (1967) Acquisition of conservation of substance through film modeling techniques. In: Brison, D. W. *and* Sullivan, E. V. (eds.), *Recent research on the acquisition of conservation of substance,* (Educational Research Series No. 2). Toronto: The Ontario Institute for Studies in Education.

SWALES, T. (1979) *Record of Personal Achievement.* Schools Council Pamphlet 16). London: Schools Council.

SWINDON EDUCATION COMMITTEE (n.d.) *Record of Personal Achievement.* Swindon: Curriculum Study and Development Centre.

THOMSON, G. H. (1950) *The Factorial Analysis of Human Ability.* London: University of London Press.

VERNON, P. E. (1950) *The Structure of Human Abilities.* London: Methuen.

WALKER, D. A. (1976) *The IEA Six Subject Survey.* John Wiley & Sons.

WATSON, J. B. (1930) *Behaviourism.* New York: W. W. Norton.

ZIGLER, E. *and* BUTTERFIELD, E. C. (1968) Motivational aspects of changes in I.Q. test performance of culturally deprived nursery school children. *Child Development,* 39, 1-14.

What happens when you collect together the most significant papers written by the retiring Director of a national educational research organisation? In the case of Bryan Dockrell — who retired from the Scottish Council for Research in Education in 1986 — you get a set of insights into matters of current concern which anyone with an interest in education or research will find stimulating and challenging.

In the papers on achievement we are asked to reflect on whether anyone really uses national surveys of attainment, and what 'intelligence' really means. Do teachers know what to do when they compare their own pupils' attainment with national norms? Do policymakers really use the data? Is intelligence 'real' or is it a figment of measurement error? These papers owe much to Bryan Dockrell's background as an educational psychologist but they are clearly informed both by his years as a teacher and by his knowledge of how decisions are made in education.

Assessment in schools is the topic most associated with his research over the last decade. His *Pupils as People* was probably the most significant book published by SCRE in the 1970s and the three papers on assessment in this volume will be of great interest to teachers and researchers alike. They cover the assessment of affective attainment, by quality, diagnostic assessment in the classroom and the views of parents, teachers and young people on what is appropriate to include in school reports. Finally we are given an overview of how assessment and reporting is dealt with in a number of countries — which leaves the impression that despite the differences, each approach seems remarkably hopeless to change.

This stimulating collection of essays is something to savour for all educationists with an interest in assessment and achievement and a lot to say to them.

SCRE Publication 97 ISBN 0 947833 25 4